NOTE TO USERS

Page(s) not included in the original manuscript and are unavailable from the author or university. The manuscript was microfilmed as received.

94

This reproduction is the best copy available.

UMI

University of Alberta

Genetic Characterization of Human *SOX* Genes

by

Heather Marie Prior

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in

Medical Sciences - Ophthalmology

Department of Ophthalmology

Edmonton, Alberta

Fall 2000

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-59655-9

Canada

University of Alberta

Library Release Form

Name of Author: Heather Marie Prior

Title of Thesis: Genetic Characterization of Human *SOX* Genes

Degree: Doctor of Philosophy

Year this Degree Granted: 2000

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Heather Prior

10424 – 28 A Avenue
Edmonton, Alberta, Canada
T6J 4J6

October 2, 2000

ABSTRACT

This research has provided novel information about the gene structure and putative functions of three human *SOX* genes, *SOX1*, *SOX2* and *SOX9*. *SOX* genes are named for a DNA-binding domain, the SRY-box, which is at least 50% identical to that of the family's defining member, the *SRY* gene (Sex determining region on the Y chromosome). Using Northern blot analysis, *SOX2* expression was observed primarily in human fetal and adult brain. Mutation screening in a patient with campomelic dysplasia, a skeletal malformation syndrome, identified a mutation in the patient's *SOX9* gene. A guanine insertion was observed at nucleotide 1456, which is predicted to disrupt the terminal 24 amino acids of the protein. Residual transactivation ability in the SOX9 mutant protein may explain why this patient survived longer than expected.

Characterization of human *SOX1* was the primary focus of this work. *SOX1* was sequenced from genomic DNA, located chromosomally, analyzed for its expression pattern, and, finally, tested in a mutation screen of patients with ocular anomalies. *SOX1* has a 1173-base pair open reading frame, with a high guanine/cytosine (GC) content of 76%, which codes for a predicted 391 amino acid protein. The DNA and protein sequences for human *SOX1* are 95% and 99% identical to the murine *Sox1* sequences, respectively. Using fluorescence *in situ*

hybridization and radiation hybrid mapping, *SOX1* was localized chromosomally to human 13q34. A widely expressed 4.6 kilobase transcript was observed using *SOX1*-specific probes on human multiple-tissue Northern blots, and a 3.9 kilobase transcript was found in fetal brain using a probe downstream of the *SOX1* open reading frame. DNA samples from patients with microphthalmia, cataracts and anterior segment disorders were tested for *SOX1* mutations by single-stranded conformation polymorphism analysis and DNA sequencing. One patient-specific *SOX1* alteration, a thymine to cytosine transition at position 1039, was found in a patient with zonular cataracts. This change may represent the first disease-causing mutations reported for *SOX1*, although it awaits confirmation by finding similar *SOX1* alterations in other cataract patients.

Taken together, these results affirm that the *SOX* genes are highly conserved transcription factors, which have significant functions in early human development.

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Genetic Characterization of Human *SOX* Genes submitted by Heather Marie Prior in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Medical Sciences - Ophthalmology.

Michael A. Walter

Ian M. MacDonald

Heather McDermid

Moira Glerum

Torben Bech-Hansen

Date: Oct 2/2000

DEDICATION

To the glory of God, whose image is indelibly inscribed on both the genetic blueprint of the human eye and the pages of my life.

ACKNOWLEDGEMENTS

Many people contributed to the work represented by this thesis. Its inception and development have been profoundly shaped by my supervisor, Dr. Michael Walter, to whom I am deeply grateful. I especially appreciated his continued faith in me as I faced the added demands of mothering while completing my degree.

My other supervisory committee members, Drs. Heather McDermid and Ian MacDonald, were also outstanding. Many thanks as well to the members of my examining committees, Drs. Ross Hodgetts, Moira Glerum, Torben Bech-Hansen, and Rachel Wevrick.

This work benefitted from the financial support of the RP Foundation – Fighting Blindness, the Medical Research Council of Canada, and the Alberta Heritage Foundation for Medical Research. I am especially grateful to the Heritage Foundation for supporting me with a full time studentship.

The daily work of research and study was facilitated and made enjoyable by every one of the members of the Ocular Genetics Research Group (Farideh, Nancy, James, Doug, Kristin, Steve, Kathy, Alan, Kerry, Sherry, Dean, Jody, Ramsey, Mai, Margaret, Sofia, and Fred). In particular I wish to thank Nancy Nesslinger, James Friedman, and Farideh Mirzayans for their remarkable helpfulness and friendship. The summer and project students I worked with were excellent without exception. Many others in the Departments of Genetics, Medical Genetics, Cytogenetics, and Ophthalmology were also a pleasure to work with.

The final words of gratitude are for my family and friends, who have stuck by me through the six years of labour on this degree. Thank you especially to Rosanne and Roger, Jonathan and Lynda, Merle and Rob, Keith and Wendy, James and Lois, David and Nancy, Darius and Nooshin, Greg, Elizabeth and Ian, Colin and Lynette, and Dr. Margaret-Ann Armour. Mom and Dad Prior and Mom and Dad Konrad have been incredibly encouraging, and the childcare provided with love by Mom and Dad Konrad allowed me the time and peace of mind I needed to complete this task. Lastly, most importantly, I thank my wonderful husband, Dean, and my son, Daniel, for their unfailing love and faith in me. Daniel's cheery disposition and affection were an ongoing source of joy and encouragement. Throughout everything, Dean gave selflessly of his time, help, love, and support. Thank you with all of my heart!! I'm looking forward to much more time spent together in the coming years.

TABLE OF CONTENTS

I. INTRODUCTION

A. Transcription factors in development 1
B. Overview of human eye development 4
C.Genetic factors in ocular development 9
D.Genetics of cataracts 22
E. SOX genes in development
1. Overview 29
2. HMG DNA-binding domain 33
3. Testis determination and spermatogenesis 35
4. Chondrogenesis (SOX5, 6, 9) 39
5. Early embryogenesis and neural development 41
6. Neural crest (SOX10) 43
7. Lens development (SOX1, 2, 3) 44
8. Hemopoiesis (SOX4) 46
9. SOX interactions 47
F. Research project rationale and description 51

II. MATERIALS AND METHODS

A. cDNA Library Screening 53
B. DNA Isolation 55
C. DNA digestion and electrophoresis 57
D. Ligation 58
E. Transformation 58
F. Random priming 59
G. Southern blots 59
H. Hybridization 60
I. RNA isolation 60
J. RNA gels and Northern blotting 61
K. PCR 62
L. Single stranded conformation polymorphism (SSCP) Analysis .. 62
M. TA cloning of PCR fragments 64
N. Sequencing 64
O. Rapid amplification of cDNA ends (RACE) 66
P. Primer extension 67
Q. Cell culture 68
R. Fluorescence *in situ* hybridization (FISH) on metaphase spreads.69
S. Radiation hybrid (RH) mapping 71

T. World wide web resources 72
U. Patient descriptions 73

III. RESULTS

A. Fetal Brain Library Screening for *SOX1* and *SOX2*
1. Library screening 76
2. Clone restriction mapping 77
3. Clone sequencing 78

B. PCR-based Strategies to Isolate *SOX1*
1. PCR amplification of *SOX1* 81
2. P1 Library screening for *SOX1* 83
3. P1 subcloning and sequencing 83

C. Localizing *SOX1*
1. Monochromosomal hybrid mapping 91
2. Fluorescence *in situ* hybridization mapping 93
3. Radiation hybrid mapping 94

D. Expression Studies
1. *SOX1* 95
2. SOX2 99

E. Transcript Mapping for *SOX1* 100
1. Primer extension 101
2. RACE and RT PCR 101
3. Genomic cloning and differential Northern blot analysis . 104
4. Fetal brain library screening 110
5. *In silico* gene feature modelling 112

F. Mutation Studies
1. *SOX1* 115
2. *SOX9* 123

IV. DISCUSSION

A. The *SOX1* gene 127
B. Localization inferences 132
C. The *SOX1* transcript: several possibilities 136
D. A new cataract gene? 142
E. Campomelic dysplasia mutation 146

V. CONCLUSIONS 149

VI. REFERENCES 152

VII. APPENDICES

A. Human *SOX* gene accession numbers 181
B. Solutions 182
C. *SOX1* Primers 185

List of Tables

Table 1. Transcription Factors Important in Eye Development
Table 2. Human Genes Mutated in Non-syndromic Cataracts
Table 3. Human SOX Genes
Table 4. Conditions for SSCP Analysis of *SOX1* and *SOX9*
Table 5. World wide web resources
Table 6. Phenotypic information for individuals tested for *SOX1* mutations
Table 7. SOX Clones Isolated by Fetal Brain cDNA Library Screening
Table 8. RACE reactions for *SOX1*
Table 9. Clone analysis from fetal brain cDNA library
Table 10. *In silico* predictions of gene features for the *SOX1* genomic region
Table 11. *SOX1* alterations detected within 1384 – MoMo2 fragment
Table 12. Phenotypes of chromosome 13q34 deletion patients

List of Figures

Figure 1. Downstream target model of transcriptional control
Figure 2. Human embryonic eye development
Figure 3. Stages of lens formation
Figure 4. 3-D model of *SRY* binding DNA
Figure 5. Graves hypothesis for sex determination by interacting SOX genes
Figure 6. *Sox1* and *Sox2* expression in the developing murine eye
Figure 7. Aberrant lens development in *Sox1* deletion mice
Figure 8. SOX-partner enhancer binding
Figure 9. Model for SOX-partner interactions
Figure 10. Restriction Maps of SOX Clones
Figure 11. PCR amplification of F/G *SOX1* fragment
Figure 12. Library RACE method for cDNA cloning
Figure 13. *Bgl II – Xho I* subcloning construct
Figure 14. Subcloning *SOX1* from P1 DNA
Figure 15. *SOX1* sequencing contig
Figure 16. *SOX1* DNA Sequence with Translation
Figure 17. Monochromosomal Hybrid Panel PCR for *SOX1* Primers 1611/2077
Figure 18. FISH localization of *SOX1* to chromosome 13q34
Figure 19. *SOX1* expression profile on Northerns with 450 bp *Pst I – Sst I* probe
Figure 20: *SOX2* expression profile on Northerns with clone B probe
Figure 21. RACE diagram
Figure 22. Subcloning of *SOX1* 3' region
Figure 23. Differential results with SOX1 probes on human fetal Northerns
Figure 24. Subcloning of *SOX1* 5' UTR
Figure 25. *SOX1* Genomic Map
Figure 26. *In silico* predictions of gene features, ESTs, and THCs for the *SOX1* genomic region
Figure 27. *SOX1* SSCP PCR contig
Figure 28. SSCP analysis for 1384 + MoMo2 fragment
Figure 29. DNA sequence showing T to C transition for patient #927
Figure 30. *Rsa I* map of the 1384 + MoMo2 PCR product
Figure 31. *Rsa I* digest of 1384 + MoMo2 PCR products for patient panel
Figure 32. *Rsa I* digest of 1384 + MoMo2 PCR products for patient #927 family
Figure 33. *SOX9* SSCP PCR contigs
Figure 34. SSCP and sequence analysis of Y + W *SOX9* product
Figure 35. Comparative structural alignment for SOX1, Sox1, SOX2, and SOX3
Figure 36. Functional domains of Sox1, 2, and 3 proteins

Figure 37. *SOX1* location on chromosome 13
Figure 38. Comparative amino acid alignment of human, mouse, and chick *SOX* genes
Figure 39. Summary of *SOX9* mutations in campomelic dysplasia

List of Symbols and Abbreviations

A, C, G, T	nucleotides (adenine, cytosine, guanine, thymine)
Amp	ampicillin
APS	ammonium persulfate
BAAD	biotinylated Anti-Avidin D
BAC	bacterial artificial chromosome
BLAST	Basic Local Alignment Search Tool
BN buffer	bicarbonate, NP-40 buffer
bp	base pairs of DNA
BSA	bovine serum albumin
cDNA	complementary DNA (DNA copy of mRNA)
C_f	final concentration
CPM	counts per minute
DAPI/PI	4',6-diamidino-2-phenylindole/propidium iodide
ddH_2O	distilled, deionized water
DEPC	diethylpyrocarbonate
dGTP	2'-deoxyguanosine-5'- triphosphate
dITP	2'-deoxyinosine-5'-triphosphate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphates
DTT	dithiothreitol
EDTA	ethylene-diamine tetra-acetic acid disodium salt
EST	expressed sequence tag
FAD	fluorescein Avidin DCS
FBS	fetal bovine serum
FCS	fetal calf serum
FISH	fluorescence *in situ* hybridization
GTB	glycerol tolerant gel buffer (Tris base,Taurine,EDTA)
GTE	glucose, Tris, EDTA solution
HMG	high mobility group
IPTG	isopropylthio-B(beta)-D-galactoside
kb	kilobases
kD	kiloDaltons
LB	Luria broth
M	molar
mRNA	messenger RNA
NT2	NTera2ND1 cell line

O/N	overnight
OG	orange G (C.I.16230; Acid Orange 10)
PAC	P1 artificial chromosome
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PE	primer extension
PHA	phytohaemagglutinin
RACE	rapid amplification of cDNA ends
RBC	red blood cells
RH	radiation hybrid
RNA	ribonucleic acid
RPE	retinal pigment epithelium
rpm	revolutions per minute
RPMI	RPMI 1640 medium
RT	reverse transcriptase
SDS	sodium dodecyl sulfate
SM (buffer)	NaCl,MgSO4,7H20, Tris.Cl,Gelatin
SOX	Sry-like HMG box
SRY	Sex-determining region on the Y chromosome
SSC	sodium chloride,sodium citrate
SSCP	single-stranded conformation polymorphism
SSPE	sodium dihydrogen orthophosphate monobasic
TBE	Tris-borate EDTA
TE	TRIS/EDTA buffer
TEMED	N-N-N'-N'-tetramethylethylenediamine
TENS	Tris-HCl, EDTA, NaOH, SDS
THC	tentative human consensus sequence
Tm	melting temperature
Tris	TRIS base
TSS	transcription start site
UTR	untranslated region
UV	ultra violet
WBC	white blood cells
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galactoside

INTRODUCTION

A. Transcription factors in development

The miracle of development transforms a single fertilized egg cell into a complex multicellular organism. This intricate process relies on a vast array of signals and interactions, both within and between cells. Ultimately, the expression of transcribed genes into a distinct set of proteins gives each cell its unique identity and function within the larger domains of tissue, organ, and organism. Thus the regulation of transcription is a fundamental basis of development control.

Vertebrate transcription by RNA polymerase II is regulated by the action of proteins, which bind to specific DNA sequences within promoters or enhancers in the context of a particular gene (Mitchell and Tjian 1989). Unlike prokaryotic genes, which tend to have small, well-defined promoters preceding the transcription initiation site, genes in higher organisms can have regulatory domains upstream, downstream, or even within the gene itself. The proteins that interact with these regulatory domains are called **transcription factors** and may activate or repress transcription, usually by means of complex contacts with other transcription factors and the basic transcriptional machinery. Key regulators of development are the genes encoding transcription factors, which control target or "downstream" gene expression. A simple model of the downstream regulation of target genes by transcription factors is depicted in **Figure 1**.

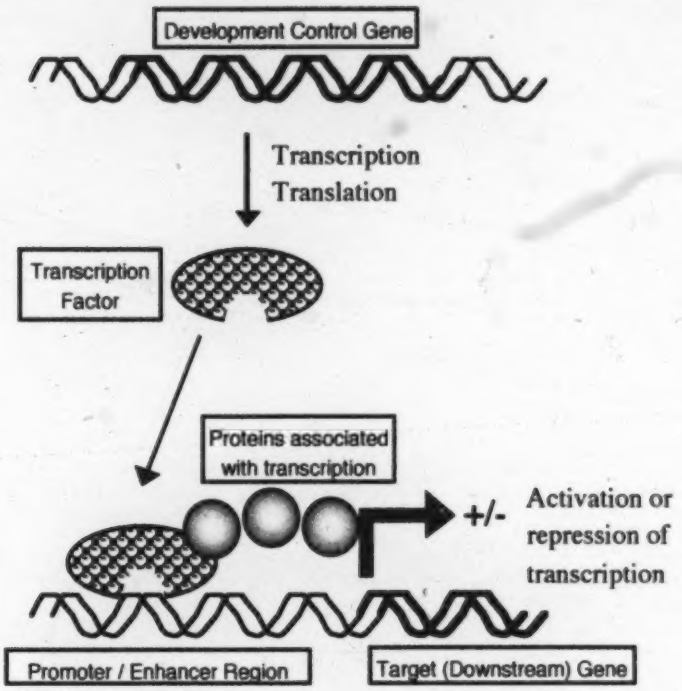


Figure 1. Downstream Target Model of Transcriptional Control

Transcription factors characteristically contain multiple domains which facilitate DNA-binding, multimerization, and/or transcriptional regulation (Tjian and Maniatis 1994). Most transcription factors are classified according to their DNA-binding domain. Although some factors have a general affinity for DNA, others bind to specific DNA sequences. Some well-studied types of sequence-specific DNA-binding domains include zinc fingers, homeodomains, leucine zippers, and paired domains (Busch and Sassone 1990).

In addition to a DNA-binding domain, classic transcription factors have distinct regulatory domains that interact with other proteins to coordinately activate or

repress gene transcription. These domains are less easily categorized than DNA-binding domains, and are usually identified according to the predominance of certain amino acids. Activation domains have been associated with regions rich in acidic residues, glutamine, proline, or serine and threonine (Triezenberg 1995; Johnson et al. 1993). Although less is known about repression domains, sequences containing alanine, glycine, proline and/or other hydrophobic residues, have been associated with transcriptional repression in *Drosophila* genes such as *Kruppel*, *engrailed*, and *Msx-1* (Han and Manley 1993; Catron et al. 1995). The fact that some domains have been found to be involved in both activation and repression highlights the ability of transcription factors to participate in extremely complex interactions with DNA and other proteins, enabling them to function in a great diversity of roles.

The discovery of large families of transcription factors has greatly enhanced our understanding of development control at the level of transcription. Many genes encoding transcription factors were first studied in *Drosophila melanogaster*, where mutants were identified by their homeotic transformation phenotype. The HOM/HOX (homeobox-containing) gene family plays a role in early pattern formation and segmentation in the *Drosophila* larva and has been implicated in the patterning of early vertebrate embryos as well (Krumlauf 1994). The human *HOXD13* gene has been found to be mutated in cases of synpolydactyly (Muragaki et al. 1996). Indeed, a growing number of transcription factors have been identified with human genetic disorders (Engelkamp and van Heyningen 1996). The *PAX* (paired box) genes are an excellent example of a developmental family that has proved to be very important in genetic disease (Hill and Hanson 1992). *PAX6* is a master eye development control gene in organisms as diverse as

insects, cephalopods, and humans. Mutations in *PAX6* cause congenital eye defects in humans, including: aniridia, Peter's anomaly, and autosomal dominant keratitis (Jordan et al. 1992; Halder et al. 1995; Mirzayans et al. 1995). Mutations in other PAX genes also cause human disease including: coloboma and renal anomalies (*PAX2*), and Waardenburg syndrome (*PAX3*) (Tassabehji et al. 1992; Sanyanusin et al. 1995). The SOX gene family is a third example of a gene family of transcription factors and is the focus of this thesis. Although SOX genes have not been as well characterized as some other gene families, they also play key roles in early development and interact with regulatory partners in complex developmental pathways.

B. Overview of human eye development

The eye is one of the human organs whose development has been studied in great detail. This detailed characterization along with the eye's relative accessibility for phenotypic observation make it an outstanding organ in which to pursue studies of the genetic control of development. Before discussing the genetic factors contributing to human eye development in more detail, a basic overview of the stages and processes involved in early eye formation is presented below in **Figure 2** (general references: Mann 1964; Barishak 1992). The early involvement of the mouse genes *Sox1*, *2*, and *3* in the formation and growth of the vertebrate lens makes them excellent genes to study when exploring the genetic control of ocular development (Kamachi et al. 1995).

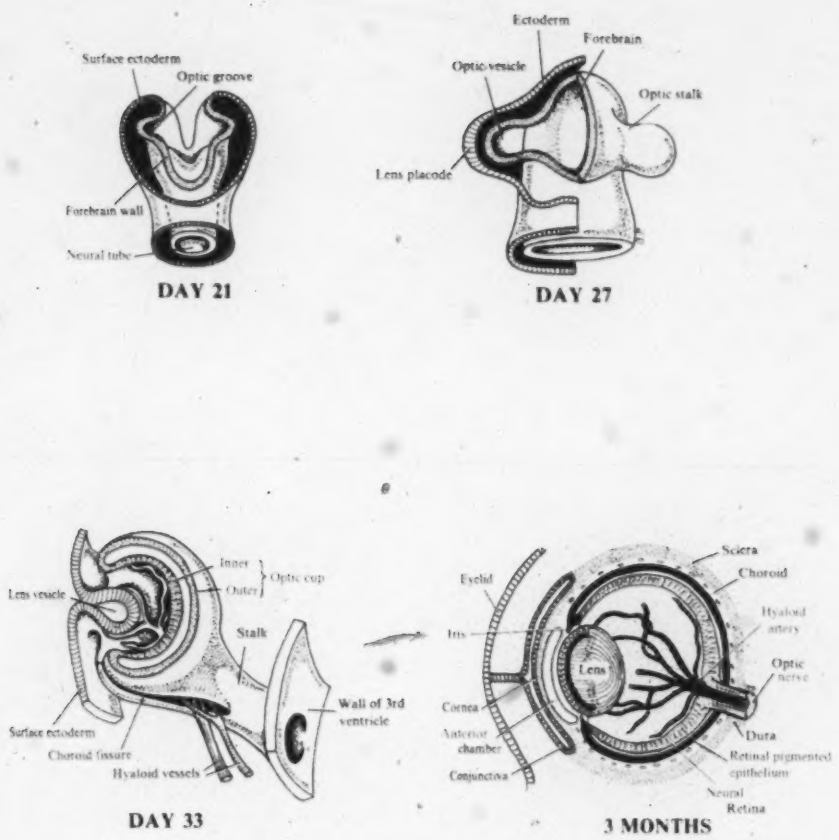


Figure 2. Human embryonic eye development (modified from Pansky 1982).

The structures that form the eye develop from several embryonic sources. The neuroectoderm of the **forebrain** gives rise to the neural retina, parts of the iris, ciliary body, and the optic nerve (structures in boldface are designated in **Figure 2**). Migrating neural crest cells, derived from the neural ectoderm, form the basis of some of the layers of the cornea and iris. The **surface ectoderm** of the head contributes primarily to lens and corneal epithelium. Finally, the mesenchyme of

the mesodermal layer between the neuroectoderm and surface ectoderm forms the vascular and schleral of the eye.

Eye development in human embryos begins early in the fourth week of gestation. At this stage, the embryo has undergone gastrulation, the formation of three germinal layers, and is in the process of neurulation, in which the rudiments of the central nervous system are laid down, culminating in the closure of the **neural tube** by the end of the fourth week. At the cephalic end of the embryo, bilateral **optic grooves** or sulci form in the neural folds, just preceding the closure of the folds to create the presumptive **forebrain**. These optic grooves grow out laterally from the lumen of the forebrain into the surrounding mesoderm and expand to create hollow **optic vesicles**. The proximal region narrows into so-called **optic stalks**, which connect the optic vesicles to the forebrain. Thus the eyes develop as outgrowths of the primitive forebrain, and have many contiguous structures with the central nervous system.

At the end of the fourth week of development, signals from the optic vesicles induce the overlying **surface ectoderm** to thicken into the **lens placodes**. Each placode invaginates, forming a lens pit, which pinches off into the mesoderm adjacent to the optic vesicle to form a **lens vesicle**. Meanwhile, the centre of the optic vesicle flattens inward and the edges elongate around the growing lens vesicle, forming a double-walled structure known as the **optic cup**.

The **choroidal fissure** forms along the ventral side of each optic cup and stalk, and invading mesenchyme leads to the development of the **hyaloid blood vessels**

in the fissure. The **hyaloid artery** and vein grow to nourish the developing lens and the inner and middle layers of the optic cup. During the sixth week, the optic fissures grow closed around the hyaloid artery and vein, enclosing them within the **optic nerve**. The distal sections of these vessels will degenerate in the latter months of gestation, but the proximal sections persist as the central vessels of the retina.

By 6 to 7 weeks of age, the human embryo has the rudiments of all of the main components of the eye. The optic cup continues to differentiate, responding to inductive signals from the lens, with the thinner outer layer developing into the **retinal pigment epithelium** (RPE) and the inner layer forming the thicker, multi-layered **neural retina**. The axons of the ganglia of the neural retina grow along the optic stalk and transform into the **optic nerve**, which continues to develop until after birth. The RPE induces surrounding mesenchymal tissue to develop into the fibrous protective covering of the eye called the **sclera**. Between the sclera and RPE, the vascular layer of the **choroid** forms, from both mesodermal and migratory neural crest progenitors.

The epithelium of the **iris** and the pigmented portion of the ciliary body develop from the anterior edge of the optic cup, once again in response to inductive signals from the lens. The musculature of the iris derives from the neuroectoderm of the optic cup, but the stroma and vascular connective tissue form from invading mesenchymal tissue of neural crest origin. The smooth muscle of the ciliary body as well as most of the trabecular meshwork are also formed from immigrating neural crest cells.

In the developing **lens**, an anterior layer of cells forms the anterior lens epithelium. In response to inductive signals from the optic cup, the cells of the posterior wall of the lens lose their nuclei and lengthen to fill the cavity of the lens vesicle as primary lens fibres. Secondary lens fibres, which also loose their nuclei and organelles to become transparent to light, are generated at the equatorial zone of the lens. Unlike primary fibres, which are only produced embryonically, secondary fibres are produced throughout life and form concentric layers around the primary fibres, causing continual slow growth of the lens. The points of contact between secondary fibers from opposite points on the equatorial zone are known as the lens sutures. **Figure 3** depicts some of the key stages of lens development.

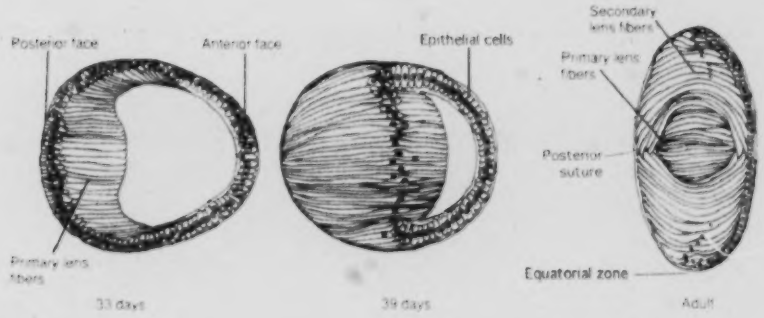


Figure 3. Stages of lens formation (modified from Corliss 1984).

Signals from the developing lens also induce the overlying surface ectoderm to give rise to the rudiments of the **cornea** and **conjunctiva**. Beneath the corneal ectoderm, the corneal endothelium and stroma develop from successive waves of immigrating neural crest cells. The **anterior chamber** then forms in the space

between the lens and the developing cornea. Ocular structures anterior to the lens are referred to collectively as the anterior segment of the eye.

C. Genetic factors in ocular development

The process of ocular development is intricately directed at every point by the action of genes. Genetic signals are exchanged between the various parts of the eye throughout its development. This interdependence was demonstrated almost a century ago in Spemann's classic series of ablation experiments on frog embryos. When the embryonic lens was removed, the retinal development was arrested. Similarly, when the retina was destroyed, lens development was halted (Spemann 1938; Grainger 1996). These experiments led to the elaboration of theories of induction, in which signals emanating from one source induce differentiation in a closely apposed region. An increasing number of genes have been identified that take part in inductive processes, in both signaling and responding to signals, as pluripotent precursors acquire competence, bias, and, finally, commitment to a certain fate.

Partly because the eye is a non-vital organ, a large number of mutations affecting it have been identified in humans and other organisms. Over a hundred distinct human ocular genetic disorders have been mapped, with a growing number boasting cloned genes (MacDonald et al. 1998). Many genes have now been identified that are expressed in the developing eye and direct its growth and differentiation. For the purposes of investigating the role of SOX genes in human eye development, this background discussion of genetic aspects of eye development will focus primarily on the ocular development genes which have been shown to play a role in human genetic disease (**Table 1**).

Table 1. Transcription Factors Important in Human Eye Development

Gene (OMIM #)	Position	Disease	Ocular Expression	References
PAX6 (106210)	11p13	Aniridia; Peter anomaly; Keratitis; Cataracts; Foveal hypoplasia; Other anomalies *Small eye* mouse	Surface ectoderm, lens vesicle, optic cup, optic stalk (6wk. embryo) Cornea, lens, retina (8-22 wk. fetus)	Jordan et al. 1992 Glaser et al. 1994 Mirzayans et al. 1995 Azuma et al. 1996 Hanson et al. 1999
PAX2 (167409)	10q25	Optic nerve coloboma and renal hypoplasia *Krd* mouse	Optic fissure and optic stalk	Sanyanusin et al. 1995
PITX2 (601542)	4q25	Axenfeld-Rieger syndrome (RIEG1) Iridogoniodysgenesis syndrome (IRID2) Peter anomaly		Semina et al. 1996 Kulak et al. 1998 Doward et al., 1999
FOXC1 (601090)	6p25	Axenfeld-Rieger syndrome Axenfeld-Rieger anomaly Iridogoniodysgenesis anomaly (IRID1) ASMD	Widespread expression of large size transcript Smaller transcripts detected in fetal colon, fetal kidney and adult leukocytes	Mears et al. 1998 Nishimira et al. 1998 Mirzayans et al. 2000
CRX (602225)	19q13.3	Leber congenital amaurosis Cone-rod dystrophy –2	Photoreceptors, pineal gland	Freund et al. 1997 Freund et al. 1998 Furukawa et al. 1997
CHX10 (142993)	14q24.3	Microphthalmia, cataracts, and iris abnormalities *Ocular retardation* mouse	Developing and mature neural retina	Burmeister et al, 1996 Percin et al. 2000
MITF (156845)	3p14-p12	Waardenburg syndrome (type II) *Microphthalmia* mouse	Melanocytes	Tassabehji et al. 1994
PAX3 (193500)	2q35	Waardenburg syndrome (types I, III) *Splotch* mouse	Neural crest derivatives	Tassabehji et al. 1992 Baldwin et al. 1992

The biggest player in the arena of genetic control of eye development is the ***PAX6*** gene. First cloned in mice and humans in the early 1990s, this gene was found to

be altered in the mouse *small eye (sey)* mutant and in human *Aniridia* patients (Walther and Gruss 1991; Hill et al. 1991; Ton et al. 1991). Subsequently, a Drosophila orthologue was identified, the *eyeless (ey)* gene, which turned out to be involved in an eye-lacking mutant known since the early 1900's (Quiring et al. 1994; Gehring and Ikeo 1999). In a remarkable series of experiments, Gehring and colleagues showed that the Drosophila *eyeless* gene was capable of directing ectopic eye formation on wings, legs and antennae when expressed in non-eye imaginal discs (Halder et al. 1995). Even more surprising was the demonstration that the mouse *Pax6* gene could perform the same function (Halder et al. 1995). These dramatic transformative and trans-phyla experiments led to the designation of *PAX6* as a master control gene for eye development and re-shaped evolutionary theory. The traditional view of diverse metazoan eyes as examples of convergent evolution (Salvini-Plawen and Mayr 1961) has been replaced by the hypothesis of an ancestral prototype eye. In this model, a photoreceptor cell and a pigment cell would have come together under the control of primitive *PAX6* to form a simple eye. Interestingly, the photosensitive molecule opsin is also conserved across phyla (Gehring and Ikeo 1999).

Recently, a second PAX6 orthologue has been identified in *Drosophila*, the *twin of eyeless* gene *(toy)* (Czerny et al. 1999). The TOY protein is more similar to mammalian PAX6 than EY in its overall sequence, including a transactivation domain which EY lacks. The two PAX6 genes, *ey* and *toy*, are believed to have arisen uniquely in holometabolous insects by an ancient gene-duplication event (Gehring and Ikeo 1999). The *toy* gene can also induce ectopic eyes in Drosophila, but requires a functional *ey* gene, suggesting that *toy* acts upstream of *ey*. Further evidence has shown that *toy* is expressed earlier than *ey*, and that *toy*

directly regulates the eye-specific enhancer of the *ey* gene, placing *toy* at the top of the eye development hierarchy (Czerny et al. 1999).

PAX6 is one of the paired-box family of transcription factors, which all contain a 128 amino acid DNA-binding domain homologous to that of the *Drosophila paired* gene (Bopp et al. 1986). In addition, various members of the *PAX* family may contain a paired-type homeodomain and/or an octapeptide domain, of which *PAX6* has a complete homeodomain but no octapeptide (Strachan and Read 1994). Human disorders have been shown to be caused by mutations in *PAX2* (Renal-coloboma syndrome), *PAX3* (Waardenburg syndrome I, III) and *PAX6* (Aniridia and other anterior segment anomalies).

All of the *PAX* genes show expression in the developing mouse embryo and play key roles in mammalian organogenesis (Mansouri et al. 1999). Expression analysis of *PAX6* by immunohistochemistry in the developing human eye has shown staining for the surface ectoderm, lens vesicle, optic cup (inner and outer layers) and optic stalk in 6 week embryos, and for the corneal epithelia and conjunctiva, lens epithelium, and non-pigmented ciliary epithelia from 8 to 22 weeks. In addition, the entire retina was positive from 8 to 10 weeks, followed by more restricted expression in the ganglia and nuclear layers after 21 weeks (Nishina et al. 1999). In spite of this widespread expression, only the prospective lens ectoderm has been shown experimentally to have an absolute requirement for PAX6 activity, which is necessary for it to maintain lens-forming competence (Quinn et al. 1996).

Knowledge of direct targets and regulators of PAX6 is still preliminary. PAX6 binding sites are found in various crystallin gene promoters. Target gene studies have demonstrated that PAX6 acts as an activator for the αB-, αA, δ1-, and ζ-crystallins, whereas it represses expression of βB1-crystallin in the developing chick lens (Duncan et al. 1998). The expression of *PAX6* in the lens epithelium, but not in the fibre cells, may function to maintain the relatively undifferentiated state of the epithelial cells by repressing βB1-crystallin and other fibre-specific genes (Kondoh 1999). PAX6 has also been shown to bind the promoter of the neural cell adhesion molecule (N-CAM) and activate its expression *in vivo* (Holst et al. 1997). However, the biological significance of this interaction is uncertain.

Complex regulatory elements in the *PAX6* promoter region direct its expression in different tissues and have been conserved across species (Xu et al. 1999). In studies of quail PAX6, binding sites for the transcription factor c-myb have been identified, and evidence has shown that PAX6 actually binds its own promoter in an autoregulatory positive feedback loop (Plaza et al. 1993).

A spectrum of human genetic eye disorders can be caused by heterozygous mutations in the *PAX6* gene. Some examples of these conditions include aniridia (Jordan et al. 1992), Peter anomaly (Hanson et al. 1994), autosomal dominant keratitis (Mirzayans et al., 1995), congenital cataracts with late-onset corneal dystrophy (Glaser et al. 1994), isolated foveal hypoplasia (Azuma et al. 1996), and ectopia pupillae (Hanson et al. 1999). A large range of mutations, along with the broad phenotypic spectrum, has made genotype-phenotype correlations difficult. Mutations are catalogued on the Human *PAX6* Allelic Variant Database

Web Site at http://www.hgu.mrc.ac.uk/Softdata/PAX6. The anterior segment malformations may be due to a disruption of the separation of the lens from the surface ectoderm, and the foveal dysplasia may result from a late-acting *PAX6* mutation, since the fovea only begins development at about 30 weeks gestation and continues until about 10 weeks after birth (Nishina et al. 1999). Considering its wide range of both temporal and spatial expression, complex interactions with other developmental factors likely greatly influence the final effect of a *PAX6* mutation on phenotype. In most cases, *PAX6* mutations are heterozygous, but in one case of a compound heterozygote carrying two mutant *PAX6* alleles, the patient had no eyes or nasal structures and severe brain malformations, a phenotype similar to homozygous *small eye* mice (Glaser et al. 1994).

One of the factors whose expression pattern overlaps that of *PAX6* is another *PAX* gene, ***PAX2***. However, unlike PAX6, the regions of the developing eye expressing PAX2 are limited to the optic vesicle, followed by the ventral optic cup and finally the optic stalk (Tellier et al. 1998). In addition to the eye and developing nervous system, a major site of PAX2 expression and function is within the developing kidney. PAX2 expression is regulated by other transcription factors; it is activated by nerve growth factor (NGF) and early growth response gene 1 (EGR1), and repressed by brain-derived neurotrophic factor (BDNF) and Willms tumour 1 (WT1) (Freund et al. 1996). Studies of key developmental signaling molecules in vertebrate eyes have shown that PAX2 and PAX6 expression are affected oppositely by members of the hedgehog family, possibly in such a way that would lead to the demarcation of retinal versus optic nerve primordia (Ekker et al. 1995).

Consistent with its expression pattern, mutations in *PAX2* have been found in patients with renal coloboma syndrome, an autosomal dominant developmental abnormality involving both hypoplastic kidneys and coloboma of the optic nerve (Sanyanusin et al. 1995). The optic nerve coloboma, which may also encompass part of the RPE, likely results from the failure of the optic fissure to close properly during development. The *Krd* mouse (kidney, retinal defect) is a mouse model of the human disorder in which the murine *PAX2* gene is deleted due to a transgene insertion (Keller et al. 1994). Similar to *PAX6*, genotype-phenotype correlations are not simple for *PAX2*. In fact, patients with the same mutation have had quite variable phenotypes, even within the same family (OMIM, entry 167409).

Two transcription factors that also play a role in human eye development and have an overlapping spectrum of mutant phenotypes are the ***PITX2*** and ***FOXC1*** genes. PITX2 is a member of the paired-like class of homeodomain factors which have a homeobox similar to that of the *Drosophila bicoid* gene (Semina et al. 1996). FOXC1 is a member of the forkhead family of transcription factors, which contain a 100 amino acid DNA binding domain homologous to the *Drosophila forkhead* gene (Pierrou et al. 1994). Mutations of either *PITX2* or *FOXC1* have been found in dominant familial cases of Axenfeld-Rieger malformation and iridogoniodysgenesis (Semina et al. 1996; Kulak et al. 1998; Mears et al. 1998; Nishimira et al. 1998; Mirzayans et al. 2000). Axenfeld-Rieger features include posterior embryotoxon (also called a prominent Schwalbe's line) and iris hypoplasia. Iridiogoniodysgenesis involves malformation of the iridocorneal angle and iris stromal hypoplasia. Both types of disorders frequently cause glaucoma. These disorders can be found as either isolated ocular disorders

(termed anomaly) or as syndromes which include maxillary hypoplasia, hypodontia, and redundant periumbilical skin.

Molecular characterization of *PITX2* and *FOXC1* is just beginning to explore the questions of their developmental expression profiles and roles in ocular organogenesis, particularly in the formation of the iridocorneal angle. In developing mice, PITX2 in situ hybridization experiments have shown expression in the periocular mesenchyme, maxillary and mandibular epithelia, and umbilicus (Semina et al. 1996). These locations are all consistent with the affected areas of Rieger syndrome. The *FOXC1* 3.9 kb transcript showed widespread expression by Northern blot analysis, and smaller transcripts were seen in fetal colon, fetal kidney and adult leukocytes (Pierrou et al. 1994).

Recent mouse models for both *PITX2* and *FOXC1* mutations have helped to further elucidate their roles in ocular development. *Foxc1* (formerly called *Mf1*) knockout mice have severe ocular anomalies, including complete absence of the anterior chamber and notably disorganized corneal layers. A pivotal role for *Foxc1* in corneal formation has been postulated from these observations (Kidson et al. 1999). *Foxc1* heterozygote mice exhibit anterior segment malformations very similar to human phenotypes, including underdeveloped Schlemm's canal, altered trabecular meshwork, iris hypoplasia, eccentric pupils and displaced Schwalbe's lines (Smith et al. 2000). In spite of these notable abnormalities, though, these *Foxc1* heterozygotes showed normal intraocular pressure and no evidence of glaucoma. The lack of similar disease outcomes in the *Foxc1* mice emphasizes the highly complex nature of eye development which is greatly

influenced by genetic background and subject to subtle differences in expression and function across species.

PITX2 has been the subject of a great deal of recent work on left-right asymmetry formation in the developing vertebrate embryo. Left-sided expression of *Pitx2* has been demonstrated in mouse and chick embryos in response to asymmetric signalling by *Nodal*, *Sonic hedgehog*, and *Lefty* (Logan et al. 1998; Yoshioka et al. 1998). The role for *PITX2* in eye development is obviously not lateralized, but the flurry of recent activity in this field has led to the production of a number of *Pitx2* mutant strains of mice with hypomorphic and null alleles. Homozygotes showed severe reduction or absence of the anterior chamber and extraocular muscles, as well as displaced pupils and disorganized cornea (Lu et al. 1999). In addition, homozygotes had optic nerve coloboma but no obvious hypoplasia of the retina (Gage et al. 1999). Heterozygotes had eyes with Rieger-like anomalies, including corectopia (asymmetrically placed pupils), anisocoria (uneven pupil size), multiple pupillary openings and clouded lenses (Gage et al. 1999). These *Pitx2* heterozygotes showed a range of severity of phenotype, reminiscent of the variability seen in Axenfeld-Rieger disease.

Another recently cloned transcription factor shown to have a role in human ocular disease is ***CRX*** (cone-rod homeobox-containing gene), a member of the *orthodenticle* group of homeodomain genes (Vandendries et al. 1996). Mutations in *CRX* underlie cases of autosomal dominant cone-rod dystrophy and Leber congenital amaurosis (Freund et al. 1997; Freund et al. 1998). Cone-rod dystrophy (CORD) is characterized by progressive loss of color vision and visual

acuity followed by night blindness and loss of peripheral vision. Some severe cases also exhibit progressive retinal pigmentation and chorioretinal atrophy (Moore 1992). Leber congenital amaurosis (LCA) is a genetically heterogeneous disorder, usually autosomal recessive, marked by severe visual loss in the first few months of life, accompanied by sharply reduced or undetectable electroretinogram and no systemic disease (Perrault et al. 1999). All *CRX* cases of CORD have been due to heterozygous *CRX* mutations. One case of LCA has been due to a homozygous mutation, as one might expect with this severe phenotype, but two LCA cases have revealed heterozygous mutations. Possibly these heterozygous LCA patients are compound heterozygotes whose second mutation has not yet been identified (OMIM entry 602225).

The 4.5 kb human CRX transcript has been detected in retina but not in ten other tissues tested (Freund et al. 1997). A more detailed analysis of murine *Crx* showed expression restricted to developing and mature photoreceptor cells. A Crx-specific binding sequence was found upstream of opsin genes as well as the gene for interphotoreceptor retinoid-binding protein. Analysis of dominant negative *Crx* mutations in photoreceptor cell culture experiments clearly indicated that the Crx transcription factor is essential for proper differentiation of photoreceptor cells (Furukawa et al. 1997). Together with conclusions drawn from mutational analysis, these studies suggest that the product of the *CRX* gene is essential in photoreceptor cells, for both early development and also ongoing maintenance.

Another gene from the paired homeobox family expressed most abundantly in the developing retina is ***CHX10*** (De Chen et al. 1989). *Chx10* null mutations are

responsible for the *ocular retardation* mouse phenotype, an autosomal recessive disorder which includes microphthalmia, thin retina, optic nerve aplasia, and developmental impairment of neural retina formation, specifically the absence of bipolar cells (Burmeister et al. 1996). Recently mutations in human *CHX10* have been found in two unrelated families with microphthalmia, cataracts, and severe abnormalities of the iris (Percin et al. 2000). Expression of *CHX10* in human eyes is found in progenitor cells of the neuroretina and the inner nuclear layer of the mature retina. The mutations identified both occurred in the homeodomain and severly disrupted binding of CHX10 to its target sequence (TAATTAGC) (Percin et al. 2000).

A final human genetic disorder caused by key development control genes is Waardenburg syndrome. This disorder is characterized by dystopia canthorum, (widely spaced inner canthi), pigment disturbances, and cochlear deafness. The ocular defects may include heterochromic irides and albinotic fundus. This syndrome shows a wide range of severity and is genetically heterogenous (OMIM entry 193500). Mutations in the transcriptions factors ***PAX3*** and ***MITF*** are responsible for cases of Waardenburg syndrome types I and III (*PAX3*), and type II (*MITF*) (Read and Newton 1997). *PAX3* is a third member of the PAX family implicated in genetic errors of development. A mouse model for Waardenburg, the *Splotch* mouse, has a disrupted *Pax3* gene (Epstein et al. 1991). The *MITF* gene encodes a basic helix-loop-helix leucine zipper protein. This transcription factor has been shown to be essential for tyrosinase expression during melanocyte development, and is mutated in the *microphthalmia* mouse, for which the gene is named (Freund et al. 1996). A third gene involved in a Waardenburg variant, the Waardenburg-Shah syndrome, is *SOX10*, which will be discussed in more detail

in the overview of SOX genes (Pingault et al. 1998). Waardenburg syndrome is an autosomal dominant trait. In cases of a homozygous affected child born to heterozygous parents, the manifestations of developmental disrupton have been much more severe, including both upper limb and neural tube defects (OMIM entry 193500).

The absence of melanocytes affects pigmentation in the skin, hair, and eyes, and hearing function in the cochlea. Therefore, disruption of melanocyte differentiation by *MITF* mutation likely causes auditory-pigmentary symptoms in Waardenburg syndrome type II. Watanabe et al. (1998) provided evidence that PAX3 transactivates the *MITF* promoter directly, and that PAX3 proteins mutant in either the paired domain or the homeodomain failed to recognize and transactivate the *MITF* promoter. These results suggested that the failure of this regulation due to PAX3 mutations causes hypopigmentation and deafness in individuals with Waardenburg syndrome types I and III. PAX3 expression has been found in neural crest cells, from which melanocytes and other neural tissues are derived. This neuronal expression pattern, along with some association of Waardenburg syndrome with neural tube defects, has implicated *PAX3* in broader neural development pathways (Chalepakis et al. 1992).

The foregoing survey of transcription factors implicated in human eye development is scarcely complete. The reader is directed to the more comprehensive review by Freund et al. (1996). In addition, a closer examination of human SOX genes in relation to development will be presented in the overview of SOX genes. Other transcription factors, including *PITX3* and *EYA1*, will be considered in the following section on the genetics of cataracts.

Several interesting themes emerge when reflecting on the role of development control genes in human disease. First, most of the disorders caused by mutations in transcription factor genes are inherited as dominant conditions. In cases when two mutant alleles are present, the phenotype is inevitably much more severe than in the heterozygote. This observation implies the highly dosage sensitive requirement for transcription factors during development. These factors must be present in precisely regulated amounts and locations in order to properly fulfill their role in the developmental cascade. Although there may be a few examples of dominant negative alleles, most mutations take effect on the basis of haploinsufficiency.

A second trend to note among these developmental disorders is the high frequency of variable phenotypes, even for similar or identical mutations. This variability likely derives from the exquisite reliance on interactions with other genes for the proper functioning of transcription factors. Modifier effects from these target or regulatory genes would thus be expected to have significant influence on phenotype.

A final observation regarding the roles of development control genes in disease is that most mutant phenotypes manifest as syndromes. This diversity of effect may reflect the various processes that transcription factors participate in within the context of different tissues. They may partner with tissue-specific cofactors to achieve unique effects during growth and differentiation. In some cases, they may act in seemingly antagonistic roles, activating or repressing transcription

depending on partnering factors. The range of syndromic defects may also reflect underlying similarities between tissues or organs which seem superficially unrelated. For instance, cochlear hair cells and iris stromal cells are both derivatives of migratory neural crest progenitors that are affected by *PAX3* mutations. Suffice it to say that when it comes to development, nothing is simple or completely isolated in its effect, and this complexity is made abundantly clear by looking at genetic aspects of eye development.

D. Genetics of cataracts

Within the field of ocular development, the maturation of the lens has long been a model system for developmental biologists. Lens growth and differentiation in vertebrates has been informative in the study of many basic developmental processes, including induction, morphogenic movements, polarization of axes, establishment of competence, programmed cell and subcellular degradation, temporal and spatial specific gene expression, and the process of aging. A highly specialized final structure and easy detection of malformation make the lens very well suited for observations of developmental abnormalities.

Several techniques have facilitated studies of lens development both *in vitro* and *in vivo*. Basic culture of lens explants was successfully performed in the 1920s (Kirby 1927), and has continued to be enhanced and exploited. Newer methods for assessing molecular aspects of development have recently been developed. Strong crystallin-specific promoters can target transgene expression selectively to the lens to observe lens-specific effects (Fini et al. 1997). Another technique for molecular analysis is an *in vivo* complementation system, in which transgenic material of interest is introduced by injection of embryonic stem cells into

blastocysts of the mouse mutant *aphakia*, which does not develop lenses. The lenses formed will reflect the transgenically introduced material (Liegeois et al. 1996). These techniques, alongside standard methods of molecular analysis, have helped to unravel the genetic control of vertebrate lens development.

Many aberrations of lens development are clinically manifest as cataracts, an opacified disturbance of the normally transparent lens. Age-related cataract is the leading cause of adult blindness worldwide, and congenital cataract is one of the most common causes of treatable childhood blindness in Europe and the USA, at a prevalence of 1.8 cases per 10 000 (Francis et al. 1999). Cataracts with a genetic etiology account for about half of congenital cataracts. Cataracts are also frequently associated with other hereditary ocular disorders, such as Norrie disease, aniridia, and retinitis pigmentosa. A range of systemic genetic disorders are also frequently associated with cataracts, a few examples of which are galactosaemia, Alport sydrome, myotonic dystrophy, Nance-Horan syndrome, and Down syndrome (Hejtmancik et al. 1995). Most isolated hereditary cataracts are autosomal dominant, and at least a dozen loci have been identified (Ionides et al. 1999). Quite a few genes mutated in isolated cataracts have been identified recently, bringing the current total to 10 (see **Table 2**).

A unified standard classification system for cataract nomenclature has not been adopted, and there are a variety of methods for categorizing cataracts, based on features such as appearance, age of onset, location, or severity. A classification used by Ionides et al. (1999) delineated eight types of autosomal dominant cataract, to which a ninth, total, can be added. These types are based on a combination of location and appearance and include: anterior polar, posterior

polar, nuclear (includes Coppock-like), lamellar (or zonular), cortical, pulverulent, coralliform, cerulean (blue-dot), and total.

Table 2. Human Genes Mutated in Non-syndromic Cataracts

Gene (OMIM#)	Protein	Location	Type of Cataract	Reference
CRYAA (123580)	α-crystallin	21q22.3	Zonular central nuclear (ADCC2)	Litt et al. 1998
CRYBA1 (123610)	β-crystallin	17q11-q12	Congenital zonular with sutural opacities (CCZS)	Kannabiran et al. 1999
CRYBB2 (123620)	β-crystallin	22q11-q12	Cerulean (CCA2)	Litt et al. 1997
CRYGC (123680)	γ-crystallin	2q33-q35	Coppock-like (CCL) Variable zonular pulverulent	Heon et al. 1999 Ren et al. 2000
CRYGD (123690)	γ-crystallin	2q33-q35	Punctate, progressive juvenile-onset Crystalline aculeiform or frosted	Stephan et al. 1999
GJA3 (121015)	Connexin46	13q11	Zonular pulverulent (CZP3)	Mackay et al. 1999
GJA8 (600897)	Connexin50	1q21.1	Zonular pulverulent (CZP1)	Shiels et al. 1998
MIP (154050)	Major intrinsic protein	12q14	Polymorphic and lamellar	Berry et al. 2000
EYA1 (601653)	Eyes-absent	8q13.3	Nuclear, with ocular anterior segment anomalies (BOR gene)	Azuma et al. 2000
PITX3 (602669)	Pitx3	10q25	Total, with or without ASMD	Semina et al. 1998

The cataract genes identified to date fall into three basic categories: crystallins, junction-associated proteins, and transcriptional regulators. These categories represent some of the fundamental molecular determinants of lens differentiation. Other molecular constituents, which will not be discussed, include growth factors, cytokines, extracellular matrix, and cytoskeletal components (Wride 1996).

A thorough molecular review of the **crystallins** has been presented by Graw (1997). These lens-defining proteins represent over ninety percent of the soluble

protein in the lens and roughly a third of the total mass of the lens. This diverse group of highly soluble, stable proteins is largely responsible for the rigid, transparent structure of the lens. Three main types of crystallins are found in the human lens. The α-crystallins are the earliest to be expressed in the lens vesicle, and are encoded by two highly similar genes, αA- and αB-crystallin, located on chromosomes 21 and 11 respectively. The β- and γ-crystallins are related globular proteins with a two domain structure containing four β-pleated sheets called 'Greek key' motifs. These crystallins are expressed most highly in the lens fibre cells, and make up the major crystallin in the lens cortex. The β-crystallin genes are found on chromosomes 17 (types A1, A3), 2 (type A2), and on 22q11-q12 (a cluster of types A4, B1, B2, B3, ψB2). The γ-crystallin genes are clustered on chromosome 2q33-q36 (types A-F, of which E and F are pseudogenes).

The crystallins are considered excellent examples of 'gene sharing,' in which a single gene encodes proteins utilized for diverse functions. The α-crystallins closely resemble small heat shock proteins and are considered to be molecular chaperones. One of the avian crystallins, δ2-crystallin, encodes the enzyme argininosuccinate lyase. Several α- and β-crystallins have been shown to have enzymatic functions as autokinases. These genes may have been co-opted for use in the lens because of their stability and solubility, resulting in diverse types of crystallin proteins. Complex regulatory elements appear to mediate lens-specific versus non-lens expression of various crystallins (Hejtmancik et al. 1995).

Cataract-causing mutations have been identified in five types of human crystallins, including αA-, βA1-, βB2-, γC-, and γD-crystallin (Table 2). The

Coppock-like cataract mutation on chromosome 2 was initially thought to be a novel pathogenic mechanism. An alteration in the promoter of the γE-crystallin pseudogene was identified that led to dramatically increased expression of the pseudogene. However, later work found the pseudogene promoter alteration among normal controls, and revealed a second, disease-associated, alteration in the γD-crystallin gene (Héon et al. 1999).

Lens transparency depends on the highly ordered array of structural elements, like the crystallins, as well as the metabolic support necessary to maintain a high protein concentration over extremely long periods of time. Mechanisms to minimize oxidative stress and to balance hydration and electrolyte concentration are vital. These needs are especially demanding within the optically clear nuclear and cortical layers of the lens in which cells have lost their nuclei and organelles. Intracellular communication is a key element in achieving this proper balance and is greatly facilitated by **junction-associated proteins**.

Gap junctions and thin junctions are two components of the extensive array of membrane channels between lens cells. Gap junctions are formed by polypeptide subunits called connexins which connect neighbouring cells via hydrophilic channels, allowing passage of ions and small molecules. Thin junctions regulate water movement between cells, and are composed primarily of two aquaporin-type proteins, major intrinsic protein (MIP) and major protein 19 (MP19) (Francis et al. 1999). Disruption in three elements of these junction systems have been shown to cause cataracts: connexin46, or gap junction protein α-3 (GJA3); connexin50, or gap junction protein α-8 (GJA8); and MIP (**Table 2**). The identification of the GJA8 gene has extra historical significance because it

represents the gene identification for the first human disease locus assigned to an autosome. The cataract locus on chromosome 1 was first linked with the Duffy blood group in 1963, and now, nearly twenty years later, the gene has been identified (Hejtmancik 1998).

Transcriptional regulators are the third type of lens component exhibiting cataract mutations. This finding is not surprising considering the importance of precise spatial and temporal control of gene expression in lens differentiation. The transcription factor *PITX3* is a *paired* type of homeodomain protein which has about 70% identity to *PITX2*, the gene implicated in Axenfeld-Rieger syndrome (Semina et al. 1998). Deletions in the promoter of murine *Pitx3* which abrogate its lens-specific expression have recently been shown to underlie the mouse mutant *aphakia (ak)* (Semina et al. 2000). In mouse embryonic expression studies, the PITX3 protein was first detected in the lens placode and invaginating lens pit, and subsequently throughout lens development. Extra-ocular expression was found in the embryonic brain, tongue, dental primordia, and mesenchymal regions of the head and upper thorax (Semina et al. 1998). Cataracts in one family with *PITX3* mutations were associated with anterior segment mesenchymal dysgenesis (ASMD), an ocular malformation involving structures derived from neural crest precursors (corneal stroma and endothelium, trabecular meshwork, iris stroma). This mutation was a 17-bp insertion near the 3' end of the gene, outside of the homeodomain, but altering a 14 amino acid motif conserved in all *PITX* genes. A second *PITX3* mutation, a G to A transition, was found upstream of the homeodomain in a mother and son with total cataracts (Semina et al. 1998).

The second transcriptional regulator involved in cataract mutation is the *EYA1* gene, a human homologue of the *Drosophila eyes absent gene (eya)* (Abdelhak et al. 1997). Flies with severe alleles of this gene are completely eyeless, and transgenic expression of *eya* can induce small ectopic eyes in the antennae and the ventral zone of the head (Desplan 1997). This gene has been described as a novel nuclear protein. Although it lacks a recognizable DNA-binding domain, Xu et al. (1997) demonstrated that the proline/serine/threonine-rich N-terminal regions of the murine Eya proteins have transcriptional activator activity. This group also showed that the mouse *Eya* gene requires *Pax6* for expression in the lens and nasal placode, suggesting that *Eya* may participate in the same developmental control pathways as *Pax6*.

Although the *eya* gene is crucial for the formation of *Drosophila* eyes, the human *EYA1* was first identified by positional cloning as the gene responsible for branchiootorenal (BOR) syndrome. This developmental malformation syndrome affects only areas of the branchial arch, ear, and kidney, and does not typically involve anomalies of the eye (OMIM entry 601653). Azuma et al. (2000) examined DNA from patients with various types of developmental eye anomalies for *EYA1* mutations. They identified three missense mutations in patients with congenital cataracts and anterior segment anomalies very reminiscent of anomalies found in cases of *PAX6* mutations. One patient also had clinical features of BOR syndrome. These findings confirmed a role for the human *EYA1* gene in eye morphogenesis, and demonstrated once again that a wide range of clinical manifestations may be caused by mutations of transcriptional regulators.

Many additional transcription factors have been implicated in human lens development. This field of investigation has blossomed recently, taking clues from model organisms as starting points, and investigating the role of highly conserved genes in human development. The master control gene *PAX6* is absolutely required for lens formation, and is likely the factor which conveys lens-forming competence to head ectoderm long before the appearance of the lens placode itself. As mentioned earlier, *PAX6* also functions as a complex regulator of crystallin gene expression (Duncan et al. 1998). A partial list of other important transcription factors for human lens differentiation includes: *AP-1*, *AP-2*, *BMP-7*, large *Mafs*, *OPTX2*, *PROX1*, *RAR*, *RXR*, *SIX-3*, *SIX-5*, *SOX1*, and *SOX2* (Freund et al. 1997; Graw 1997; Konhoh 1999). A specific discussion of the roles of *SOX1* and *SOX2* in the regulation of crystallin gene expression will be given within the context of background information on *SOX* genes.

E. *SOX* genes in development

1. Overview

The development control genes which are the focus of this thesis are the ***SOX*** genes, SRY-related HMG box genes. *SOX* genes represent a growing gene family that has already been shown to have diverse roles in human development (reviewed in Wegner 1999; Penvy and Lovell-Badge 1997; Prior and Walter 1996). The *SOX* genes as a group were initially identified by their homology to the DNA-binding domain of *SRY* (Sex-determining Region on the Y chromosome), the mammalian testis-determining gene (Denny et al. 1992). In addition to possessing a DNA-binding domain, numerous *SOX* genes have demonstrated transcriptional activation activity, affirming that this is a family of transcription factors important in development.

SOX genes have been found to be highly conserved in many species. *SOX* genes in lower organisms include the *Sox70D/Dichaete/Fish-hook* gene important for *Drosophila* segmentation and neural development (Nambu and Nambu 1996; Russell et al. 1996), and the *C. elegans COG-2* gene, implicated in vulval-uterine formation (Hanna-Rose et al. 1999). *SOX* genes have also been found to have diverse but conserved functions in vertebrates, including fish, amphibians, reptiles, birds, and mammals (Vriz and Lovell-Badge 1995; Miyata et al. 1996; Uwanogho et al. 1995; Coriat et al. 1993; Foster and Graves 1994).

Although most initial work was done on mouse *Sox* genes, a growing number of human *SOX* genes have been characterized. About 30 *SOX* genes have been recognized and grouped according to the similarities of their SRY box regions (groups A through H, **Table 3**). Apart from *SRY* on the Y chromosome and *SOX3* on the X chromosome, the other human *SOX* genes are autosomal and scattered throughout the genome. Human *SOX* genes belonging to groups A, B, and C are believed to be single exon genes, whereas those in groups D through H have a small number of introns (Wegner 1999). The *Drosophila* and *C. elegans Sox* genes can be classed with group B. Although there appears to be a predominance of observed *SOX* gene expression in the developing testis and nervous system, further study is needed to determine whether this expression pattern may have functional or evolutionary significance. The following sections will present a thematic overview of the roles of *SOX* genes in various aspects of development.

Table 3. Human *SOX* Genes

Group	Gene	Location	Size	Expression Pattern	Comments and References
A	*SRY*	Yp11	204 aa no introns in ORF 1.1 kb transcript pseudogene (2.5 kb upstream)	Male-specific gonadal ridge adult testis (Sertoli cells)	XY sex reversal No direct target identified PDZ1-binding domain(C-term) Binds AACAAT Gubbay et al. 1990; Sinclair et al. 1990
B	*SOX1*	13q34	391aa no introns in ORF 3.9kb and 4.6kb transcripts	Fetal brain 3.9 kb, 4.6 kb transcripts Most other fetal and adult tissues 4.6 kb Highest in fetal kidney	Malas et al. 1997; Pevny et al. 1998; this report
	SOX2	3q26-q27	317 aa no introns in ORF 3.5 kb transcript in NT2 cells	NT2 cells Embryonic CNS	Stevanovic et al. 1994
	SOX3	Xq27.1	443 aa no introns in ORF 2.3 kb transcript	RT-PCR Expression: Fetal brain, spinal cord, adrenals, liver, thymus, spleen, pancreas Adult liver, spleen, heart Embryonic CNS	Collignon et al. 1996; Mumm et al. 1997; Stevanovic et al. 1993
	SOX14	3q23	240 aa no introns in ORF 1.8 kb transcript (Arsic) 2.5kb transcript (Malas)	Liver cell line on Northern RT-PCR Expression: Fetal brain, spinal cord, thymus Weak other	Proline-rich Close to loci for blepharophimosis, ptosis, and epicanthus inversus syndrome and Moebius syndrome Arsic et al. 1998; Malas et al. 1999; Hargrave et al. 2000
	SOX21	13q31-q32	276 aa 5 kb transcript	Embryonic brain	Malas et al. 1999
C	*SOX4*	6p23	474 aa no introns in ORF 5.2 kb transcript Minor 3.9 kb transcri in testis	Lymphocytes Melanocytes Adult testis (3.9kb)	Able to transactivate CD2 gene Farr at al. 1993
	SOX11	2p25	441aa no introns in ORF 9.5kb,3.0kb transcrip	Fetal brain, lung, kidney No adult tissues Embryonic CNS(early) and PNS (late)	Jay et al. 1995
	*Sox12** *Sox19**				*Wright et al. 1993*
	SOX22	20p13	315 aa no introns in ORF 3.5 kb transcript 5 kb minor transcript 1.5 kb heart, skeletal muscle transcript	Widespread expression Fetal brain, kidney Adult heart, pancreas, testis, ovary Highest in embryonic CNS	Jay et al. 1997

Group	Gene	Location	Size	Expression Pattern	Comments and References
D	*SOX5*	12p12.1 8q21(ψ)	347 aa At least 5 introns Alternate splicing 1.8 kb adult testis transcript larger fetal brain transcripts	RT-PCR Expression: Widespread Northern expression: Fetal brain, Adult testis	Wunderle et al. 1996
	*L-Sox5**		*679 aa(cf 392 aa for Sox5)*	*Mouse sites of chondrogenesis*	*N-terminal coiled domain mediates dimerization* *Lefebvre et al. 1998*
	*Sox6**			*Mouse sites of chondrogenesis* *Mouse developing CNS*	*N-terminal coiled domain mediates dimerization* *Connor et al. 1995*
	SOX13	1q32	890aa likely has introns	Kidney, lung, liver	N-term leucine zipper Roose et al. 1999
E	*SOX8*	16p13.3	447aa three exons	N/I	2 Transactivation domains Binds WWCAAWG Pfeiffer et al. 2000
	SOX9	17q24	509 aa 2 introns 4.5 kb transcript	Embryonic testes, Sertoli cells, ovary, kidney Fetal brain Adult testis, heart	Campomelic dysplasia, Autosomal sex reversal C-term108aa TA domain Activates COL2A1 Binds AACAAT, AACAAAG Wagner et al. 1994; Foster et al. 1994
	SOX10	22q13.1	466 aa 3introns 2.9 kb transcript 3.1 kb(Bondurand) 2.7kb in testis	Fetal brain Adult heart, small intestine, colon, CNS Neural crest derivatives	TA domain C-term 90aa Waardenburg-Hirschprung disease Kuhlbrodt et al. 1998b; Pusch et al. 1998; Pingault et al. 1998
F	*Sox7** *Sox17**				*Kanai et al. 1996*
	SOX18	20q13.33	384 aa 1.9 kb transcript	Adult heart Weakly in brain, liver, testis leukocytes	Azuma T et al. 2000
G	*SOX12* *Sox15** *Sox16**	N/I	N/I	N/I	Goze et al. 1993
	SOX20	17p13	233 aa one intron after aa17 1.5 kb transcript	Northern expression: Exclusively fetal testis RT-PCR Expression: Fibroblasts, lymphoblasts	Interacts with AACAAT Hiraoka et al. 1998; Meyer et al. 1996
H	*SOX30*	5q33	753aa 501aa (skipped exon) (3 kb transcript in mouse)	Adult testes only	Osaki et al. 1999

*Human *SOX* genes are listed except in cases where no human orthologue has been cloned, in which case the murine *Sox* gene is given. Note that human *SOX12* and murine *Sox12* are not orthologous.

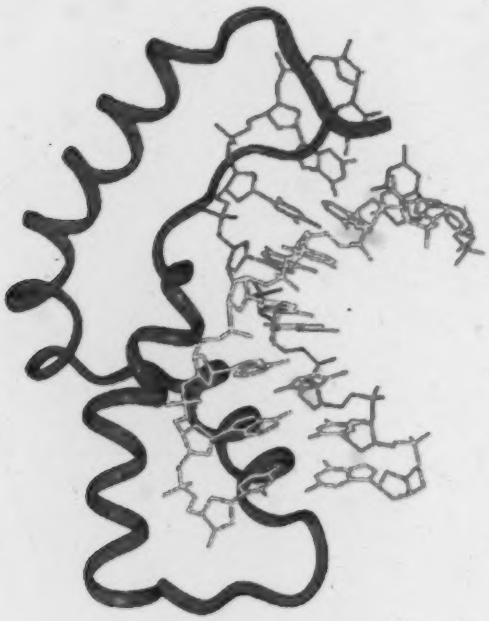
Abbreviations: N/I not indicated, aa amino acid, CNS central nervous system, ψ pseudogene

2. HMG DNA-binding domain

The *SOX* genes belong to a large group of genes whose DNA-binding domain is called a High Mobility Group (HMG) box (Laudet et al. 1993). The *SOX* genes have an HMG box which is at least 60% similar or 50% identical to the 79 amino acid HMG box of the *SRY* gene. HMG class proteins fall into two basic configurations. One type contains multiple HMG boxes and has a general affinity for binding DNA independent of sequence. This group includes HMG-1 protein, Ubiquitous Binding Factor (UBF), and Mitochondrial Transcription Factor 1 (MT-TF1). Members of the second category of HMG class proteins contain a single HMG box that binds DNA in a highly sequence-specific manner. This group includes the yeast mating type genes *matMc* and *mat*-A1, the white cell regulatory genes *T-cell Factor-1* (*TCF-1*) and *Lymphocyte Enhancer Factor-1* (*LEF-1*) (Waterman and Jones 1990; Travis et al. 1991), as well as the *SOX* genes, including *SRY*.

The sequence-specific HMG box proteins, which include the *SOX* genes, *TCF-1*, and others, have about 25% identity over the 79 amino acid HMG box, with little or no similarity outside this box. Unlike most other DNA-binding proteins, they bind DNA in the minor groove at a consensus binding sequence of A/T A/T C A A A G (Harley et al. 1994). The structure of a complex between a DNA octamer and the DNA-binding domain of SRY shows the molecular basis of this binding specificity (**Figure 4**; Werner et al. 1995). The HMG box of SRY forms an L-shape, composed of three α-helices and an extended amino acid stretch at the N-terminus. The inner surface of SRY contacts the minor groove of the DNA over quite a broad surface area.

Figure 4. 3D model of *SRY* HMG domain binding DNA (adapted from Bianchi and Beltrame 1998)



Legend:
This model depicts the complex between the HMG box of human *SRY* and the octanucleotide GCACAAAC. The backbone of the HMG-box (solid) is composed of three a-helices and an extended N-terminus. The DNA (outlined) is significantly distorted: the minor groove is widened to accommodate the extended stretch of the HMG box; the bases are tilted; and the double helix is unwound and bent by 70° – 80°. This model is derived from the atomic structure work of Werner et al. (1995).

SRY and other sequence-specific HMG class proteins induce a sharp bend of approximately 80° in the DNA template upon binding (Ferrariet al. 1992; Giese et al. 1992). Binding of the protein causes significant widening of the minor groove to accommodate the extended segment of the HMG domain, and the planes of the bases are tilted, resulting in the double helix unwinding and bending back by 70°–80°. This bending mechanism may bring different regulatory regions of the target gene into closer proximity with one another. Various transcription factors bound to these regions may then interact with one another to activate

transcription. DNA-binding proteins with this structural mode of action have been appropriately called architectural transcription factors (Grosschedl et al. 1994).

3. Testis determination and spermatogenesis (*SRY, SOX5, SOX9, SOX30*)

The first *SOX* gene to be successfully cloned and characterized was the *SRY* gene, and it remains the defining member of the SOX family (Gubbay et al. 1990). The critical region for testis determination on the Y chromosome was initially established by analyzing the DNA of sex reversed patients, including XX males with some Y material translocated onto the X chromosome, and XY females with deletions of the Y chromosome. In spite of other promising candidates, such as the Zinc Finger on the Y chromosome gene (*ZFY*), *SRY* was finally shown to be the necessary gene for male sex determination in 1990, when mutations within its open reading frame were found in XY sex reversed patients (Sinclair et al. 1990; Berta et al. 1990). These mutations all fell within the HMG box region of *SRY*, presumably disrupting its ability to bind DNA. Later experiments demonstrated that this gene was sufficient for male sex determination since XX transgenic mice carrying the *SRY* gene were phenotypically male (Koopman et al. 1991).

The SRY protein is expressed during human gonadal development in the bipotential genital ridge of the developing 46XY embryo. As development continues, expression becomes localized to the nuclei of Sertoli cells (Hanley et al. 2000). Outside of the genital ridge, *SRY* mRNA levels are regulated by an unusual translational control system in which a unique splicing event produces non-functional circular transcripts (Capel et al. 1993). In the absence of *SRY*, the

gonad will develop as an ovary in the default female pathway. In male development, Sertoli cells secrete AMH (anti-Müllerian hormone), also known as MIS (Müllerian Inhibiting Substance), which causes regression of the Müllerian duct system (female) and progression of the Wolffian ducts (male) and gonadal development as testis.

The human *SRY* gene has no identified activation domain, and nearly all sex-reversing *SRY* mutations fall within the DNA-binding HMG box. These observations suggest that the binding and concomitant bending may be the critical facets of SRY function (Pontiggia et al. 1994). SRY may mediate transcriptional regulation via interaction with an intervening factor that does contain a regulatory domain. Support for this idea comes from the results of a yeast two-hybrid screen which identified a PDZ protein called SIP-1 (SRY interacting protein-1) as an interacting protein with human SRY. Biochemical analysis and expression of SIP-1 in human embryonic testis supported the proposal that the two proteins could interact together. The interaction required the C-terminal seven amino acids of SRY and the PDZ domains of SIP-1 (Poulat et al. 1997). A SOX-binding consensus sequence in the AMH promoter seemed like a natural site of SRY action. However, affinity studies to determine the precise downstream targets of SRY have not shown that AMH or other components of the male steroidogenesis pathway are direct targets, leaving the conclusion that sex determination is not simply under the direction of one gene alone (Haqq et al. 1994).

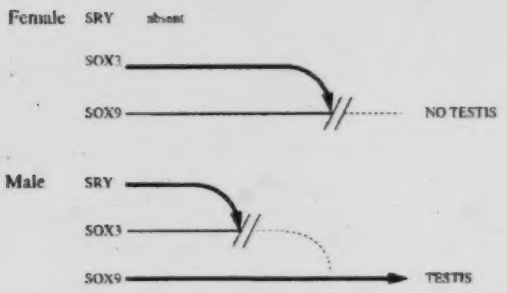
Recent evidence suggests that a second *SOX* gene, *SOX9*, also plays an important role in human sex determination (Koopman 1999). The *SOX9* gene originally

gained recognition for causing the skeletal malformation syndrome, campomelic dysplasia, when disrupted (Foster et al. 1994; Wagner et al. 1994). The observation that about two-thirds of XY patients who have this syndrome are sex-reversed females immediately suggested a role for *SOX9* in sex determination. Expression studies in developing human gonads have revealed high levels of SOX9 in presumptive testes, in a similar pattern to SRY, and only later and low level expression in developing ovaries (Hanley et al. 2000). Intriguing results of binding and transactivation studies have strongly suggested that SOX9 and another protein called steroidogenic factor 1 (SF-1) synergistically activate AMH expression in developing Sertoli cells (De Santa Barbara et al. 1998).

If SOX9 and SF-1 are the factors for activating AMH expression, what is the role for SRY? A detailed discussion of human sex determination mechanisms and models is beyond the scope of this thesis. However, an interesting hypothesis from the point of view of *SOX* genes in development has been put forward by Jennifer Graves (1998). She suggests a basic pathway in which *SRY* inhibits another gene, *SOX3*, whose product in turn inhibits *SOX9* (**Figure 5**) (see also McElreavey et al. 1993). Thus the presence of *SRY* would indirectly allow the action of SOX9 to proceed and male differentiation to occur. In the absence of *SRY*, an inhibitory action by SOX3 would prevent SOX9 expression, and development would proceed along the default ovarian pathway. The proposed roles of *SRY* and *SOX9* are well-supported by experimental evidence. *SOX3* has the highest similarity to the HMG box of *SRY*, and may be the ancestral homologue of *SRY* (Collignon et al. 1996). In humans, *SOX3* shows widespread expression in fetal tissues, including brain and spinal cord, as well as in some adult tissues, including testis (Stevanovic et al. 1993). However, there is currently

little direct evidence for a role for *SOX3* in sex determination. Other genes known to be involved in sex determination, such as *DAX1*, may prove to act as the inhibitor which Graves proposes (Swain et al. 1998).

Figure 5. Graves hypothesis for control of sex determination by *SOX* genes (Graves 1998)



Sox5 and *Sox6* are both expressed in adult mouse testis. In the mouse, *Sox5* is exclusively expressed in post-meiotic round spermatids, and may play a role in directing the process of spermatogenesis (Denny et al. 1992). The human *SOX5* transcript, by contrast, is expressed by RT-PCR assay in many tissues, most strongly in fetal brain (about 9.5 kb) and adult testis (about 1.8 kb, like the mouse) by Northern blot (Wunderle et al. 1996). The murine Sox17 protein is found in pre-meiotic spermatogonia (Kanai et al. 1996), and the murine orthologue of the recently identified human *SOX30* gene has been shown to have germ cell-specific expression in testis, making these two good candidates for involvement in spermatogenesis (Osaki et al. 1999). Another prime candidate for testis-specific function is human *SOX20*, which was found to be expressed exclusively in fetal testis (Hiraoka et al. 1998).

4. Chondrogenesis (*SOX5, SOX6, SOX9*)

Aside from its role in testis development, *SOX9* also has a vital function in chondrogenesis. Heterozygous mutations in this gene can cause campomelic dysplasia (CD), with or without autosomal sex reversal (Foster et al. 1994; Wagner et al. 1994). CD is a rare congenital skeletal malformation syndrome characterized by bowing of the long bones and defects in cartilage formation. Hypoplastic scapulae, pelvic abnormalities, bowing of the tibiae and femora with resulting pretibial dimpling, and 11 pairs of ribs, along with a variety of nonskeletal abnormalities are characteristic of CD (Houston et al. 1983, Kwok et al. 1995; Meyer et al. 1997; Hageman et al. 1998). It is associated with autosomal sex reversal, and two-thirds of XY CD patients develop with female or ambiguous genitalia (Houston et al. 1983; McKusick 1992). Death frequently occurs neonatally as a result of respiratory insufficiencies, but life expectancy varies widely, depending on the severity of phenotype (Houston et al. 1983; Mansour et al. 1995). Patients with chromosome translocations were used to localize the syndrome to distal 17q (Tommerup et al. 1993). *SOX9* was investigated as a candidate gene because mouse *Sox9* mapped to the homologous region and had been shown to have a primary role in skeletal formation (Wright et al. 1993; Wright et al. 1995). Although *SOX9* mutations have been detected in a majority of non-translocation patients, several CD translocation breakpoints have been mapped well outside of the *SOX9* gene (Wirth et al. 1996). Transgene experiments have shown that at least some of these external breakpoints cause deletions of *SOX9* regulatory elements (Wunderle et al. 1998).

The distal C-terminus of SOX9 has been identified as a transactivating domain, and most CD mutations disrupt either the HMG DNA-binding domain or the C-terminus (Südbeck et al. 1996). *SOX9* acts as a transcription factor that is known to have a role in the expression of anti-Müllerian hormone, discussed earlier, and *COL2A1*, a major collagen gene (Bell et al. 1997; De Santa Barbara et al. 1998). Detailed studies of mouse chondrogenesis have shown *Sox9* is expressed in the mesenchymal condensations that initiate skeletal formation, both in the prechondrogenic precursors and later in the maturing chondrocytes (Ng et al. 1997; Wright et al. 1995). *Sox9* binds a chondrocyte-specific enhancer in the first intron of the *Col2A1* gene and can activate *Col2A1* expression in tissue culture experiments (Lefebvre et al. 1997). Alterations in the enhancer that prevent *Sox9* binding abolish *Col2A1* expression, indicating that *Col2A1* is a direct target of *Sox9* (Bell et al. 1997).

Recent evidence has suggested that *Sox9* is not the only *SOX* gene involved in regulation of the expression of chondrogenic factors. Four HMG-type target sequences were identified in a 48 bp enhancer of the *Col2A1* gene, three of which were required for cartilage-specific expression *in vivo* in mice (Zhou et al. 1998). An alternate longer form of *Sox5* (*L-Sox5*), *Sox6*, and *Sox9* were shown to be co-expressed in chondrocytes and to act in a cooperative manner to stimulate the highest levels of expression of *Col2A1* as well as another chondrocytic marker, *aggrecan* (Lefebvre et al. 1998). *SOX* gene interaction, as well as alternative transcript findings, will be revisited further in a later discussion.

5. Early embryogenesis and neural development (*SOX* B and C Groups)

Preliminary work suggests a role for *SOX* genes in the early events of embryogenesis. In the pre-gastrulation and early somite stages of mouse development, *Sox1*, *Sox2*, and *Sox3* transcripts have all been detected (Wood and Episkopou 1999). *Sox2* and *Sox3* are expressed earliest, first in the epiblast and then becoming restricted to the prospective neural plate. They are also expressed in the primitive streak ectoderm, gut endoderm, and prospective sensory placodes. *Sox1* is first detected in the neural fold ectoderm, and all three genes are expressed in the neuroectoderm during somitogenesis. A vital role for *Sox2* has been inferred from the fact that mouse embryos deleted for *Sox2* die around implantation (Pevny et al. 1998). As well, *Sox2* has been shown to act synergistically with another transcription factor, *Oct-3/4*, to direct even earlier expression of *fibroblast growth factor 4* (*Fgf4*) , *osteopontin*, and *undifferentiated embryonic transcription factor 1 (UTF1)* in embryonic stem cells (Yuan et al. 1995; Botquin et al. 1998; Nishimoto et al. 1999). In general, *Sox2* seems to be a factor important in maintaining the undifferentiated state of early embryonic precursors.

Numerous *SOX* genes show expression in the developing nervous system, although precise roles for most of these genes have not been established. The best-studied participants in neuronal development are members of Group B. *Sox1*, *Sox2*, and *Sox3* are all expressed at high levels in the murine embryonic CNS (Collignon et al. 1996). Detailed studies of *Sox1* in neurogenesis in the mouse have shown that it is an early marker of neural fate, and *Sox1* expression can trigger neural fate decision in the embryonic P19 cell line (Pevny et al. 1998).

Chicken *Sox2* and *Sox3* are expressed in the undifferentiated cells of the neural epithelium, but then are down-regulated in maturing neurons (Uwanogho et al. 1995; Rex et al, 1997a). Recent studies in frog embryos have suggested that *Sox2* is a pan-neural marker, essential for conveying neural identity to early neuroectoderm cells (Mizuseki et al. 1998; Kishi et al. 2000). As *Sox1*, *Sox2*, and *Sox3* expression appears to be lost as cells become terminally differentiated, these genes likely play an important part in defining the neural competence for dividing precursors of the embryonic CNS. Two other closely related group B members, *Sox14* and *Sox21*, show positionally restricted expression patterns in the chicken CNS, suggesting a possible role in dorsoventral patterning of that part of the nervous system (Rex et al. 1997b; Uchikawa et al. 1999).

Several group C genes also show specific CNS expression. Chicken and mouse Sox11 proteins are upregulated in developing neurons, suggesting a role in neuronal maturation (Uwanogho et al. 1995; Hargrave et al. 1997). The human *SOX11* gene was found to be expressed in the primitive CNS, and then in the peripheral nervous system in 6-7 week embryos and throughout the brain of a 19 week fetus (Jay et al. 1995). Another group C gene, *SOX22*, is expressed most abundantly in human embryonic CNS, but not in the surface ectoderm, prompting speculation that this gene may serve as a switch during differentiation of the primitive neuroectoderm into surface ectoderm (Jay et al. 1997).

In spite of extensive evidence of expression in the developing nervous system, no neurological disorders have yet been ascribed to *SOX* gene mutations. The best candidate disorder remains the X-linked mental retardation syndrome Börjeson-Forsman-Lehmann syndrome, whose critical region overlaps the location of *SOX3*

on chromosome Xq27.1 (Stevanovic et al. 1993; Mumm et al. 1997). Animal knockout models would help to uncover prospective phenotypes for mutant *SOX* genes, but few *Sox* knockouts have been reported to date. In addition to the *Sox2* deletion mouse which did not survive past implantation, a *Sox1* deletion mouse showed a mild CNS phenotype in the form of spontaneous seizures, as well as cataracts and microphthalmia (Nishiguchi et al. 1998).

6. Neural crest (*SOX10*)

Another *SOX* gene with a unique expression profile in the developing nervous system is *SOX10*. Mutations in this group E gene have been recently shown to underlie cases of autosomal dominant Waardenburg-Shah syndrome, a disorder characterized by the pigmentary-hearing defects of Waardenburg syndrome in association with Hirschsprung disease (aganglionic megacolon) (Kuhlbrodt et al. 1998b; Pingault et al. 1998). Waardenburg-Shah patients with *SOX10* mutations have a fairly broad range of phenotype. One mild case of a Yemenite deaf-blind patient is due to a supposedly conservative missense *SOX10* mutation (Bondurand et al. 1999), and three Waardenburg-Shah cases with an associated neurologic deficit are due to a particular *SOX10* truncating mutation (Touraine et al. 2000). A spontaneous mouse mutant, *Dominant megacolon (Dom)*,which has very a similar phenotype to Waardenburg-Shah syndrome, also carries a mutation in its *Sox10* gene (Southard-Smith et al. 1998; Herbarth et al. 1998).

The Waardenburg-Shah syndrome is regarded as a neurocristopathy. *SOX10* has been found to be expressed during human embryonic development within the neural crest and its derivatives (Bondurand et al. 1998). Its expression begins in

the neural crest cells, and then becomes focused in the neural crest lineages that contribute to the peripheral nervous system, including: sympathetic, sensory, and enteric ganglia. Expression has also been found in melanoblasts, which likely cause the pigmentary effects in Waardenburg syndrome. Molecular evidence has shown that *Sox10* functions as a transcriptional activator in glial cells. It directly activates the *Protein zero* gene, one of the myelin-forming genes expressed exclusively in the Schwann cell lineage of the peripheral nervous system (Peirano et al. 2000). In mice, *Sox10* expression has also been characterized in the central nervous system, where it gradually increases in intensity until reaching a maximum in adult oligodendrocytic glia (Kuhlbrodt et al. 1998a).

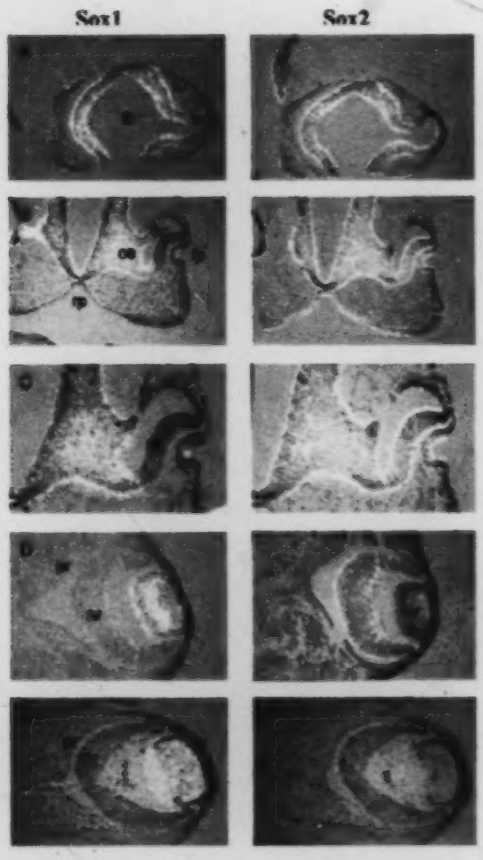
7. Lens development (*SOX1, SOX2, SOX3*)

In another part of the nervous system, a highly specific role for *SOX* genes has been identified in the developing lens. Studies of crystallin gene expression in chick lens have shown that *Sox1*, *Sox2*, and *Sox3* are the three constituents of a complex called δEF2, which binds the δ-crystallin promoter and the DC5 enhancer located in the gene's third intron (Kamachi and Kondoh 1993; Kamachi et al. 1995; Kamachi et al. 1998). This binding is essential for promoter activity, but only effects lens-specific activation of δ-crystallin expression (Kamachi et al. 1999). SOX proteins have likewise been shown to bind the γ-crystallin promoter in mice, which express γ-crystallin instead of δ-crystallin (Nishiguchi et al. 1998).

The expression pattern in the developing mouse lens is shown is **Figure 6** for *Sox1* and *Sox2*. Unlike the chick, *Sox3* does not show significant expression in the developing mouse lens (Collignon et al. 1996). *Sox2* is expressed at highest

levels in the surface ectoderm of the presumptive lens placode, and then down-regulated once lens vesicle formation occurs, giving way to the rising expression of *Sox1* in lens epithelial and elongating fibre cells (Kamachi et al. 1998).

Figure 6. *Sox1* and *Sox2* expression in the developing murine eye (from Kamachi et al. 1998)



Legend

White areas indicate *in situ* hybridization of *Sox1* and *Sox2* probes.

(A,B) Frontal sections through the head of 9.5 and 10.5 dpc embryos, respectively.

(C) Higher magnification of B. In the lateral head ectoderm Sox2 expression becomes restricted to and augmented in the area overlying the optic vesicle and forming the lens placode. As the lens invaginates, Sox1 begins to be expressed at low levels in the lens vesicle.

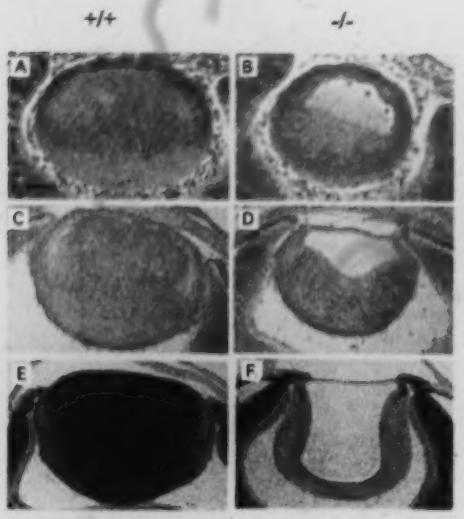
(D,E) Transverse sections of 11.5 and 13.5 dpc embryos, respectively. As Sox1 expression progressively increases, Sox2 is down-regulated.

Abbreviations: dpc, days post coitim; fb, forebrain; ov, optic vesicle; os, optic stalk; oc, optic cup; lp, lens pit; pr, presumptive pigmented retina; nr, presumptive neural retina; lv, lens vesicle; l, lens; rp, Rathke's pouch.

The major phenotype in homozygous *Sox1* deletion mice is small eyes with cataracts, and closer study has shown that they have impaired lens fibre elongation and severely reduced γ-crystallin expression (**Figure 7**; Nishiguchi et al. 1998). A small amount of two types of γ-crystallin is produced just preceding

the rise in *Sox1* expression, but this expression is likely due to activation by *Sox2*, which basically ceases once *Sox1* expression begins. These *Sox1* deletion mice express α- and β-crystallins at normal levels, suggesting that they are regulated by other factors. The heterozygotes had no detectable anomalies.

Figure 7. Aberrant lens development in *Sox1* deletion mice (adapted from Nishiguchi et al. 1998)



Legend:
Histological study of wild type (+/+) and *Sox1* deletion (-/-) lenses. Preparations were made at 12.5 dpc (A, B), 15.5 dpc (C, D), and postnatal day zero (E, F). Bar in frame A indicates 50 μm in A and B; 100 μm in C and D; 140 μm in E and F.

8. Hemopoiesis and cardiac development (*SOX4*)

One of the few *SOX* genes shown to interact with non-SOX HMG proteins is *SOX4*. Studies in developing murine lymphoblasts have shown that *Sox4* is expressed in T-cells and pre-B lymphocytes and is involved together with *TCF-1* and *LEF-1* in controlling lymphocyte differentiation (van de Wetering et al. 1993). Binding studies have shown that the Sox4 protein has affinity for the T-cell enhancer motif AACAAAG. Sox4 also has a serine-rich transactivation

domain separable from its DNA-binding domain, and was the first *SOX* gene shown to have the characteristic structure of a classical transcription factor.

An important role for *Sox4* in cardiac development can be inferred from the phenotype of a targeted *Sox4* deletion mouse. These mice suffered from embryonic lethality due to impaired development of endocardial ridges and subsequent failure in heart formation (Schilham et al. 1996). Closer study of the B- and T-cell lineages in these mice showed that B-cell development was blocked at an early stage. T-cell development was only slightly impaired, indicating that, although expressed in both types of lymphocytes, *Sox4* is more crucial for B-cell development than T-cell development (Schilham et al. 1997). Another *Sox* candidate for cardiac involvement is the mouse *Sox6* gene. Mice with inversions involving this gene exhibit myopathy, heart block, and sudden neonatal death. *Sox6* is expressed abundantly in skeletal muscle tissue, but further studies are required to establish a convincing link between *Sox6* and heart development (Hagiwara et al. 2000).

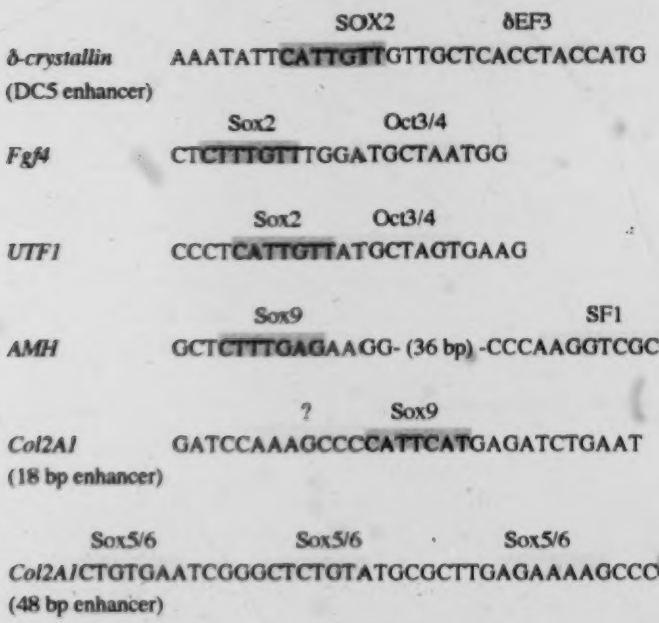
9. SOX interactions

Several investigations show the interactive mode of action that *SOX* genes adopt with one another and with other factors. Most SOX proteins studied to date bind identical or very comparable sequences and effect similar bending of target DNA. These similarities raise the question of how SOX factors achieve highly tissue-specific action. The answer may lie in the partnering of SOX proteins with other factors that are themselves present in limited domains (Bianchi and Beltrame 1998; Kamachi et al. 2000).

Sox2 binds the DC5 enhancer for the δ-crystallin gene of developing chick lens. Sox2 activation of the enhancer only occurs when another lens-specific factor, δEF3, is bound to an adjacent site (Kamachi et al. 1995). As mentioned earlier, Sox2 activates embryonic stem cell expression of *fibroblast growth factor 4* (*Fgf4*) when bound next to a second transcription factor, Oct3/4, in the *Fgf4* minimal enhancer (Yuan et al. 1995). Similar tandem binding of Sox2 and Oct3/4 is required for expression of *UTF1*, a stem cell-specific co-activator (Nishimoto et al. 1999).

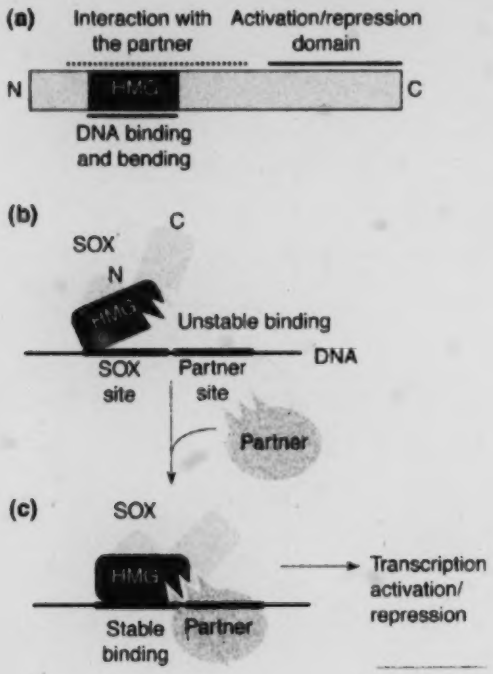
Partnerships have also been demonstrated for *Sox9* in both chondrogenesis and male sex development. Two studies have shown that Sox9 activation of enhancers within the collagen gene *Col2A1* is dependent on co-factor binding. In one case, the partners have been identified as the L-Sox5/Sox6 pair (Lefebvre et al. 1997; Lefebrve et al. 1998). In developing Sertoli cells in the male gonad, Sox9 has been shown to cooperate with the orphan nuclear receptor SF1 to activate expression of the *anti-Müllerian hormone* (*AMH*) gene (de Santa Barbara et al. 1998). **Figure 8** illustrates the experimentally determined binding of various *Sox* genes in relation to their partner sites within the aforementioned enhancers.

Figure 8. SOX-partner enhancer binding (adapted from Kamachi et al. 2000; Lefebvre et al. 1997; Lefebrve et al. 1998)



Other partnerships have been suggested for *Sox* genes, including *Sox10* and *Sox11* with two POU transcription factors (*Oct6* and *Brn1/Brn2* respectively) in Schwann cells and oligodendrocytes (Kuhlbrodt et al. 1998). Even SRY has a proposed partner to effect trans-regulation, the SIP-1 protein (Poulat et al. 1997). Detailed dissections of the binding and activation requirements for *Sox1/2/3* in comparison with *Sox9* have shown that the pairing of *SOX* genes is mediated by gene-specific interacting regions (Kamachi et al. 1999). Although DNA-binding and activation domains may be interchangeable, partner selectivity is unique to each *SOX* gene. The limited availability of the partners dictates specific times and places of action. A model for this mode of interaction is depicted in **Figure 9**.

Figure 9. Model for SOX-partner interactions (adapted from Kamachi et al. 2000)



The research into *SOX* genes is still in a relatively early stage. Most studies to date have focused on isolating and sequencing various members of the family, along with expression studies and DNA-binding affinities. The genes which have been studied in more depth are those with a known function, particularly the disease-associated members, *SRY*, *SOX9*, and *SOX10*. In surveying the current state of *SOX* science, some common features are beginning to emerge. Further study will clarify whether these trends are true for most *SOX* genes or just a few.

Currently, human *SOX* genes are attractive positional candidates for various diseases, and further mutation analyses will help to correlate structural domains

with function. Because *SOX* genes have been found in diverse organisms, model systems for studying human mutation and disease have been developed in other species, particularly knockout mice, for specific *SOX* genes. The work on *SOX* genes is just beginning. Future studies of these architects of development will certainly do much to shape our understanding of the complex process of growth and development.

F. Research project description and rationale

My research plan focused on the study of human *SOX* genes on a molecular level. My objective was to characterize these genes and their products, which are excellent candidates for involvement in hereditary developmental disorders. *SOX1* and *SOX2* were chosen because of their likely role in early neuronal and eye development. In the mouse, *Sox1*, *Sox2*, and *Sox3* are all expressed dynamically in the embryonic eye and central nervous system (Collignon et al., 1996). Therefore these genes have a potential role in human embryogenesis, and in eye development in particular. As I was interested in *SOX* genes as a whole, when an opportunity for a mutation study on *SOX9* arose, I added *SOX9* to my list of *SOX* genes under study.

Since more work had already been published about *SOX2* than *SOX1*, I focused much of my efforts on *SOX1* specifically. My research goals were as follows:

1. Obtain human cDNA clones for *SOX1* and *SOX2*.
2. Restriction map and sequence the clones.
3. Cytogenetically map the *SOX1* gene by multiple mapping methods.

4. Correlate the chromosomal location of *SOX1* and *SOX2* with loci of known human genetic disorders.

5. Screen patients with likely disorders for mutations and polymorphisms in *SOX1* and *SOX9* genes.

6. Transcript map *SOX1* as a basis for further functional analysis.

7. Determine the tissue-specific expression pattern for *SOX1* and *SOX2* as a clue to their function.

MATERIALS AND METHODS

Refer to **Appendix B** for solution compositions. A list of symbols and abbreviations follows the thesis table of contents. Refer to Sambrook et al. 1989 for a general molecular biology techniques reference.

A. cDNA Library Screening

1. Preparing plating bacteria
    a. Streak out plating bacteria (XL1-Blue) from frozen stocks onto LB plates onto which tetracycline has been spread (10 µL of 5 mg/mL); grow O/N at 37°C for isolated colonies.
    b. Pick one colony and culture in 50 mL of LB + 0.2% maltose + 10 mM $MgSO_4$ in a sterile erlenmeyer flask with shaking O/N at 30°C.
    c. Spin culture in two 50 mL conical tubes for 10 min. at 2000 rpm. Discard supernatant.
    d. Adjust OD_{600} to 0.5 with 10 mM $MgSO_4$. Store plating bacteria at 4°C for 1 – 2 weeks.

2. Titering phage library
    a. Vortex library and prepare serial dilutions in SM buffer.
    b. Mix 100 µL of each dilution of phage with 200 µL of plating bacteria in 5 mL culture tubes. As a control, include one tube without phage, and one with neither phage nor bacteria.
    c. Incubate 15 min. at 37°C.
    d. Add 4 mL top agar (preheated to 50°C), invert to mix, and quickly pour onto prewarmed LB plates (100 mm round).
    e. Allow plates to set 10 min., then incubate at 37°C O/N.
    f. Count plaques to determine library titre. The titre determined for the first fetal brain cDNA library screening was 4.8×10^6 pfu/µL, and for the second fetal brain cDNA library was 4.5×10^6 pfu/µL.

3. Plating the library
    a. Calculate desired number of plaques per plate, and prepare serial dilutions of phage in SM accordingly. Approximately 50 000 plaques per 200 mm square plate were used for both cDNA library screenings.
    b. Vortex phage. Mix 100 µL of appropriate library dilution with 1.5 mL of plating bacteria in 50 mL conical tube.

c. Incubate15 min. at 37°C.

d. Add approximately 35 mL of top agar (preheated to 50°C), invert to mix, and quickly pour onto prewarmed large LB plates (200 mm square).

e. Allow plates to set 20 min., then incubate at 37°C for 12 hrs.

4. Lifting the library

a. Chill plates for several hours after incubation.

b. Gently set precut nylon membranes (Hybond™) onto plates for 2 min. Mark orientation on membrane by pricking asymmetrically with small needle dipped in India ink.

c. Transfer membrane to denaturing solution in shallow tray, DNA-side up, for 2 min.

d. Transfer membrane to neutralizing solution in shallow tray, DNA-side up, for 2 min.

e. Rinse in 2xSSC for 2 min., then dry on filter paper.

f. Crosslink DNA to membrane with auto crosslinker (12000mJ).

5. Probing the library

a. Prehybridize the membranes in 10 mL of Church and Gilbert hybridization solution (Church and Gilbert 1984) per bottle for 2 or more hours at 65°C in rotating oven.

b. Label the probe and remove unincorporated nucleotides (see below for labeling and cleaning methods).

c. Boil the probe for 5 min. to denature, and then put on ice immediately. Add directly to hybridization solution and return bottles to oven to hybridize while rotating O/N at 65°C.

d. Wash the membranes with two quick room temperature rinses in 2xSSC, 2 x 15 min. in 2xSSC, 0.1% SDS at 50°C, and 2 x 20 min in 0.2xSSC, 0.1% SDS at 65°C.

e. Autoradiograph to detect signal.

6. Picking plaques

a. With the autoradiograph still overlying the membranes, position plates over autoradiograph by aligning ink markings on membranes with plates.

b. Using the large end of a glass pipette, excise the plaques corresponding to strong positive signals on the film, and place each agar plug into 1 mL of SM with one drop of chloroform added. Vortex.

7. Secondary and tertiary screenings
   a. Titre secondary phage stocks as done for the primary stocks. Mix phage with plating bacteria and plate desired number onto 100 mm LB plates. I used 100 – 300 plaques per plate to obtain isoloated plaques. Proceed to culture, lift, and probe as with primary phage stock.
   b. Align plates over autoradiograph and pick tertiary stock using narrow end of glass pipette. Place into 1 mL SM with one drop of chloroform, vortex.
   c. Repeat procedure for tertiary screening. Pick well-isolated plaques into 1 mL SM with one drop chloroform. Store at 4°C.

8. Excision protocol
   a. Combine 200 µL plating bacteria, 250 µL phage stock, and 1 µL of ExAssist helper phage (Stratagene) in 15 mL culture tube.
   b. Incubate 15 min. at 37°C.
   c. Add 3 mL LB to each tube. Incubate O/N at 37°C with shaking.
   d. Heat tubes at 70°C for 20 min. Spin.
   e. Decant supernatant (containing phage) into clean tube.
   f. Mix 100 mL phage supernatant with 200 mL SOLR cells (Stratagene).
   g. Incubate 15 min. at 37°C.
   h. Plate 100 mL of sample onto LB/Amp plates and incubate O/N at 37°C.
   i. Pick isolated colonies, culture O/N at 37°C, miniprep and analyze DNA.

B. DNA Isolation

1. Plasmid DNA 10 minute miniprep method
   a. Inoculate 3 – 5 mL of LB/Amp (100µg/µL) with isolated colony of interest. Grow O/N at 37°C with shaking.
   b. Spin culture in 1.5 mL eppendorf tubes. Shake out supernatant. Vortex vigourously to resuspend pellet completely in residual media.
   c. Add 300 µL fresh TENS buffer. Vortex briefly.
   d. Add 150 µL 3 M NaOAc (ph 5.2). Vortex briefly. Spin. 5 min.
   e. Remove supernatant to fresh tube. Fill tube with ice cold 95% ethanol.
   f. Spin for 20 min. at 4°C.

g. Remove ethanol. Add 100 µL of 70% cold ethanol to wash. Remove ethanol and allow to dry.

h. Resuspend in 50 – 100 µL of ddH_2O or TE/RNase. Store at -20°C.

2. Plasmid or gel slice DNA with Qiagen spin columns: follow manufacturer's instructions. Elute DNA in 50 µL ddH_2O.

3. P1 DNA

a. Inoculate 3 mL of LB/Kanomycin (25 µg/mL) with frozen P1 stock. Grow O/N at 37°C with shaking.

b. Inoculate 75 mL of LB/Kanomycin (25 µg/mL) in a 125 mL erlenmeyer flask with 2.5 mL of O/N culture. Grow 1.5 hrs. at 37°C with shaking.

c. Add IPTG to 0.5 mM (375 µL of 100 mM stock). Grow 5 hrs at 37°C.

d. Divide culture into 6 x 10 mL aliquots in 15 mL culture tubes. Spin for 10 min. at 10000 rpm or 15 min. at 2000 rpm. Discard supernatant. Pellets may be frozen at this point.

e. Resuspend 3 pellets in 1 mL of GTE per tube with pipetting. Add 30 µL of 50 mg/mL lysozyme per tube. Swirl and incubate 5 min. at room temperature.

f. Add 2 mL of 0.2N NaOH, 1% SDS. Cap firmly. Invert to mix. Place on ice for 5 min.

g. Add 1.5 mL of 3M KAc. Cap, invert, place on ice for 5 min.

h. Spin 10 min at 10000 rpm or 15 min. at 4000 rpm. Remove supernatant with pipette to 15 mL conical polypropylene tube. May be placed at 4°C if needed.

i. Add 20 µL of RNase A (final concentration 50 µg/mL). Incubate 30 min. at 37°C.

j. Add equal volume phenol/chloroform (about 4 mL total). Mix 10 times by inversion. Spin for 2 min. at 5000 rpm.

k. Remove aqueous layer to new 15 mL culture tube containing 4 mL of isopropanol. Mix by inversion. Spin 30 Min. at 4000 rpm at 4°C.

l. Pour off alcohol. Add 600 µL of 70% ethanol. Spin briefly. Remove alcohol and allow to dry in hood for about 30 – 40 min.

m. Resuspend pellet in 100 µL of ddH_2O. Transfer to 1.5 mL microfuge tube. Store DNA at -20°C

4. Lymphocyte DNA

a. Collect blood samples in EDTA tubes. Spin 10 min. at 2000 rpm at 4°C.

b. Discard plasma and RBC layer. Transfer WBC layer (buffy coat) to 50 mL conical tube.

c. Fill tube to 30 mL with warm RBC lysis buffer. Place at 37 °C for 3 – 4 min. until RBC are lysed.

d. Spin 10 min. at 2000 rpm at 4°C. Discard supernatant.

e. Add 20 mL of 0.15 M NaCl. Bang tube to dissolve pellet. Spin 10 min. at 2000 rpm at 4°C. Discard supernatant.

f. Repeat NaCl wash (step "e").

g. Add 2 mL high TE buffer. Dissolve pellet.

h. Add 2 ml WBC lysis buffer at 37°C. Swirl to mix. Place at 4°C O/N.

i. Extract with equal volumes of organic solvents: twice with phenol and twice with chloroform: isoamyl alcohol (24:1).

j. Precipitate DNA by adding 300 µL NaOAc (1/10 volume) and 7.5 mL cold 95% ethanol (2.5 volumes) and fishing out DNA pellet with pipette tip.

k. Rinse pellet by dipping in 70% ethanol, and air dry.

l. Resuspend pellet in 200 – 400 µL of ddH_2O or TE.

C. DNA digestion and electrophoresis

1. Combine reagents, adjusting the amounts of DNA and water to utilize approximately 1 µg of plasmid DNA:

    10 µL DNA
    5 µL H2O
    2 µL BSA
    2 µL reaction buffer
    1 µL restriction enzyme
    20 µL total

2. Incubate reactions at appropriate temperature for a minimum of 2 hours.
3. Prepare 0.5 – 2.0 % agarose gel with TBE buffer and 2% volume of 10 mg/mL ethidium bromide.
4. Add loading dye such as OG or xylene cyanol bromophenol blue to DNA reactions (10% by volume). Mix and load into wells on gel.
5. Apply electrical field and monitor progress of migration visually by observing position of the dye band. Photograph on UV light source when finished.

D. Ligation

1. Digest vector and insert DNA with appropriate restriction enzymes to generate compatible ends. A general ratio of 3 molar ends of insert to vector is a good starting point for most ligation reactions.
2. Combine vector and insert DNA and remove digestion buffers by combining the samples and precipitating them together.
    a. Add 10% volume 3M NaOAC + 2.5 volumes of ice cold 95% ethanol.
    b. Mix, then spin for 20 min. at 4°C.
    c. Remove ethanol. Add 100 µL of 70% cold ethanol to wash. Remove ethanol and allow to dry.
3. Set up ligation reaction by adding the following to the vector/insert pellet:
    15 µL H2O
    4 µL 5x T4 ligase buffer
    1 µL T4 ligase
    20 µL total
4. Incubate reaction at 15°C for 20 hrs. for blunt end ligations, or at room temperature for 4 – 6 hrs. for sticky end ligations.

E. Transformation

1. Thaw 50 – 200 µL of competent cells per reaction. Ice 5 min.
2. Add 1 – 10 µL of ligated DNA to cells. Ice 15 min.
3. Heat pulse for 90 seconds by placing in 42°C water bath.
4. Place tubes on ice. Add 800 – 900 µL cold LB.
5. Incubate 45 min. at 37°C.
6. Plate 100 µL per reaction onto LB/Amp plates which have been prepared by spreading with 40 µL each of 100 mM IPTG and 20mg/mL X-Gal.
7. Incubate O/N at 37°C for 16 – 20 hrs. Place at 4°C for several hours before picking white colonies for O/N culture for miniprep DNA isolation (above method B1).

F. Random priming

1. Combine probe DNA and H_2O. Boil 10 min. Place on ice for 5 min.
   5 µL DNA (about 100 ng)
   18 µL H_2O
2. Add the following reagents, on ice:
   2 µL of each non-labeled nucleotide (A, G, T)
   15 µL random primer buffer
   1 µL Klenow enzyme
   5 µL of ^{32}P-labelled dCTP
   50 µL total
3. Incubate reaction at 37°C for at least 2 hours.
4. Remove unincorporated nucleotides using a Qiagen column following the manufacturer's instructions. Elute probe in 200 µL of ddH_2O, and count 2 µL for specific activity. An average activity of $1 - 2 \times 10^6$ CPM/mL was used.
5. Calculation of specific activity for 10 mL of hybridization solution:

$$\square \frac{\text{CPM}}{2\ \mu\text{L}} \times 200\ \mu\text{L} \times \frac{1}{10\ \text{mL}} = \square \frac{\text{CPM}}{\text{mL}}$$

G. Southern blots

1. Run agarose gel with DNA samples. Cut off lower left corner of gel and photograph adjacent to ruler.
2. Denature gel by gently shaking in denaturing solution for 30 – 60 min.
3. Neutralize gel by gently shaking in neutralizing solution for 30 – 60 min.
4. Place gel top-side down onto glass plate covered with wet filter paper wick making direct contact with ample 10xSSC solution in dish below.
5. Place precut, prewetted nylon blotting membrane (Hybond™, Amersham) on top of gel. Smooth out bubbles by rolling a pipette over membrane (never touch membrane with ungloved fingers).
6. Stack several precut, prewet filter papers on top of membrane, followed by a 5 – 10 cm stack of paper towelling. Place a half-full 100 mL glass bottle for weight on top.
7. Transfer O/N. Remove membrane, dry, crosslink in auto crosslinker.

H. Hybridization

1. Prehybridize membranes by placing in hybridization bottles with 10 mL of hybridization solution (Church and Gilbert 1984) for several hours at 65°C in the rotating hybridization oven. Alternative hybridization solution used was ExpressHyb (Clontech).
2. Boil labelled probe for 5 min. to denature. Place on ice immediately.
3. Add probe directly to hybridization solution (fresh solution may be used, but it must be prewarmed).
4. Allow hybridization to proceed O/N at 65°C in rotating hybridization oven or for at least one hour if using ExpressHyb.
5. Discard hybridization solution appropriately, rinse membranes in 2xSSC, and carry out washes, either in the hybridization bottles in the rotating hybridization oven, or in plastic dishes on a shaker. Washes for Southerns and Northerns were generally performed as follows:
    a. Southerns
        2x15 min. at 50 °C in 2xSSC, 0.1% SDS
        2x20 min. at 65 °C in 0.2xSSC, 0.1% SDS
    b. Northerns
        2x10 min. at room temp. in 2xSSC, 0.05% SDS
        2x20 min. at 50 °C in 0.1xSSC, 0.1% SDS

I. RNA isolation

1. From tissues (including: frozen mouse eyes, testis, seminal vesicles, ovary, fresh human lenses)
    a. Homogenize tissue in dounce homogenizer with 1 mL of Trizol™ reagent until fully pulverized. Let stand 5 min. at room temp. Transfer to 1.5 mL eppendorf tubes
    b. Add 200 µL chloroform to tube, cap tightly, and shake vigourously by hand. Let stand at room temp. for 3 min.
    c. Spin for 15 min. at 12000 g at 4°C. Transfer aqueous phase to fresh tube and add 0.5 mL isopropanol. Incubate 10 min. at room temp.
    d. Spin for 10 min. at 12000 g at 4°C. Discard supernatant. Wash pellet by adding 1 mL of 70% ethanol, vortexing, and spinning for 5 min. at 12000 g at 4°C.
    e. Remove ethanol, dry pellet for 5 – 10 min., dissolve in 50 µL DEPC H_2O. Store RNA at -80°C.

2. From NTera2 cells

a. Discard media from cultured cells by pipette and add 18 mL of Trizol™ reagent to culture dish. Pipette up and down, and then allow to stand 5 min. at room temp. Transfer to 50 mL conical tube.

b. Add 3.6 mL chloroform to tube, cap tightly, and shake vigourously by hand. Let stand at room temp. for 3 min.

c. Spin for 15 min. at 4200 rpm at 4°C. Transfer aqueous phase to fresh 50 mL conical tube and add 9 mL isopropanol. Mix.

d. Spin for 10 min. at 4200 rpm at 4°C. Discard supernatant. Wash pellet by adding 18 mL of 70% ethanol, vortexing, and spinning for 5 min. at 4200 rpm at 4°C.

e. Remove ethanol, dry pellet for 5 – 10 min., dissolve in 100 – 200 µL DEPC H_2O. Store RNA at -80°C.

J. RNA gels and Northern blots

1. To prepare 1.2% 100 mL gel for RNA:

a. Clean equipment with 0.1 N NaOH or RNase-ZAP™ to eliminate RNases.

b. Boil 1.2 g agarose with 84.6 mL DEPC H_2O, and then cool to 65°C.

c. Add 10 mL of 10xMOPS and 5.4 mL of formaldehyde (37%w/v) in fume hood, swirl to mix and pour into clean gel tray.

2. Sample and RNA-ladder preparation:

a. Mix components according to the following proportions:

6.8 µL sample (or ladder) RNA + H_2O
2.0 µL 10xMOPS
1.2 µL formaldehyde
10.0 µL formamide
20 µL total

b. Heat samples at 55°C for 10 min. Chill on ice.

c. Add load dye (0.5 µL ethidium bromide and 3.0 µL marker dye) and load samples into wells.

d. Run gel at 120 V for 2 hrs. Photograph gel alongside ruler.

e. Northern blotting: rinse gel in DEPC-treated H_2O for 2 x 20 min., then transfer O/N exactly as for Southern blot (described above).

f. After transfer, mark lane and ladder locations, dry blot, and crosslink with UV crosslinker.

K. PCR

1. Templates: 10 – 100 ng sample DNA, positive control (genomic) DNA, H_2O
2. Primers: 100 ng each (see **Appendix C** for sequences)
   Most primers were designed using the web-based Primer 3 design program at http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi.
3. Reaction for single sample:
   a. Mix components at room temperature:
      2.5 μL 10x PCR Buffer
      2.5 μL 2 mM dNTP mix
      0.2 μL primer 1
      0.2 μL primer 2
      0.125 μL 100xBSA
   b. Add DNA (normally 10 μL) and 1 drop oil.
   c. Hot start by heating reaction to 95°C and holding.
   d. Add 0.1 units *Taq* polymerase, and H_2O to final volume of 25 μL, and begin cycling.
4. Cycling: Typical PCR reactions are carried out starting with 5 min. at 95°C, followed by 30 cycles of 30 sec. at 95°C, 30 sec. at 60°C, 30 sec. at 72°C, and followed by 5 min. at 72°C.
5. Variations between reactions included annealing temperatures, annealing or extension times, number of cycles, and amounts of DMSO (added to the reactions to obtain greater stringency). Reactions were often carried out in a 12.5 μL volume, but 25 μL and 50 μL reactions were also used.
6. After adding load dye, products are run on agarose gels for sizing.

L. Single stranded conformation polymorphism (SSCP) Analysis

1. Reactions: PCR reactions for SSCP analysis are performed like PCR reactions described above, except that dATP is reduced to 0.25 mM in the dNTP mix, and 0.5 μL (0.5 μCi) of ^{35}S-labeled dATP is added to each reaction. After cycling, 6 μL of stop dye is added to each reaction.
2. Gel preparation: SSCP PCR reactions are run on non-denaturing polyacrylamide gels with varying amounts of glycerol (from 0 to 7%). A standard 7% glycerol gel is prepared with the following reagents:

6.3 mL glycerol
9 mL 10xTBE
13.5 mL acrylamide/bisacrylamide (19:1)
61.2 mL H2O
90 mL total
Add 450 μL of fresh 10% APS and 45 μL of TEMED. Pour immediately. Set for 1 hr.

3. Electrophoresis: After denaturing for 5 min. at 95°C, 6 μL of each sample is loaded per lane and run at 60W at 4 °C. Running times range from 2 to 12 hours, depending on product size, glycerol concentration, and temperature.
4. Specific conditions for SSCP analysis of *SOX1* and *SOX9* are given in **Table 4**.

Table 4: Conditions for SSCP Analysis of *SOX1* and *SOX9*

Gene, Exon	Primer Pair	Fragment Size (bp)	PCR Conditions* (°C, % DMSO)	PAGE Conditions (% glycerol, run time)
SOX1	355 + 631	275	55, 5%	0%, 3 hr
	582 + 857	276	60, 5%	0%, 3 hr
	820 + 1181	362	62, 5%	7%, 7 hr
	1043 + 1241	189	60, 5%	0%, 3 hr
	1180 + 1473	304	56, 5%*	7%, 7 hr
	1384 + MoMo2	355	58, 5%	7%, 7 hr
***SOX9*, Exon 1**	int1 + 410	139	70, 0%	7%, 3 hr
	507 + B	156	62, 0%	7%, 3 hr
	696 + int2nn	364	65, 0%	7%, 8 hr
***SOX9*, Exon 2**	int3 + G	185	65, 0%	7%, 3 hr
	277 + int4	238	65, 0%	7%, 7 hr
***SOX9*, Exon 3**	int5 + Q	124	65, 0%	7%, 5 hr
	337 + 693	168	70, 0%	7%, 3.5 hr
	M + 692	183	65, 0%	7%, 3.5 hr
	773 + 691	203	61, 7%	7%, 6 hr
	U + X	252	67, 0%	7%, 7 hr
	Y + W	322	67, 0%	7%, 7 hr

* cycling times: normally denature 30s, anneal 30s, extend 30s ; for *SOX1* primer pair 1180 + 1473, denature 45s, anneal 30s, extend 60s

M. TA cloning of PCR fragments

1. Cut desired PCR product band out of agarose gel. Isolate DNA with Qiagen spin column according to instructions. Resuspend DNA in 45 µL ddH_20.
2. To add A's, combine the following reagents, top with one drop oil, and incubate for 20 min. at 72°C:

    45 µL PCR product DNA isolated from gel
    5 µL 10x PCR buffer
    0.2 µL Taq polymerase
    50.2 µL total
3. Purify product by adding to 250 µL PB reagent (Qiagen). Run onto Qiagen spin column, then rinse column in PE reagent (Qiagen), and elute in 30 µL ddH2O.
4. Ligate PCR product into vector by combining the following and incubating O/N at 4°C:

    3 µL A-added, purified DNA
    5 µL 2xLigation buffer (Promega)
    1 µL pGEM-T Easy vector (Promega)
    1 µL T4 ligase
    10 µL total
5. Transform TA ligation product into competent cells with standard procedure (see section E).

N. DNA Sequencing

1. Manual sequencing:

In order to overcome the difficulty of sequencing regions with high GC content, I developed a protocol with slight modifications to standard sequencing reactions using the Amersham ^{33}P Thermosequenase sequencing kit (Amersham Life Sciences). These modifications included using less template DNA and primer, adding DMSO, doubling the amount of enzyme used, using dITP instead of dGTP nucleotide mixtures, decreasing the proportion of ddNTP to dNTP, adding a preliminary denaturing step before cycling, and using an increased number of cycles. All of the *SOX1* open reading frame sequence was determined using manual sequencing. Reactions were carried out as follows:

a. Combine reagents in the following proportions:

5 μL template DNA (100 – 500 ng)
2 μL primer (5 – 10 ng)
5 μL H_2O
1 μL DMSO
2 μL reaction buffer
4 μL Thermosequenase enzyme
19 μL total

b. Aliquot 4.3 μL of reaction mixture to four tubes, GATC.

c. Prepare nucleotide mixture:

10 μL dITP nucleotide master mix
1.0 μL 33PddNTP

d. Aliquot 2.5 μL of nucleotide mixture to the four reaction tubes, GATC.

e. Cylcing conditions:

95°C 5 min.
50 cycles of:
95°C 30 sec.
55°C 20 sec.
60°C 8 min.

f. Add 4 μL stop dye to each reaction. Prepare a 6% polyacrylamide gel with the following reagents:

40.5 g urea
40 mL ddH_2O
4.5 mL 20xGTB
13 mL acrylamide/bisacrylamide (19:1)
90 mL total
Filter and degas solution under vacuum. Add 900 μL of fresh 10% APS and 23 μL of TEMED. Pour immediately. Set for 1 hr.

g. Denature sequencing products for 2 min. at 85°C before loading 3.5 μL per lane. Run sequencing products at room temp. at 50W for 2 – 8 hrs. Transfer gel to large Whatman filter paper, dry under vacuum at 80°C for 2.5 hrs., and autoradiograph at room temp., overnight or longer.

h. As an alternative to the dITP protocol given above, some reactions were carried out using dGTP, with a higher annealing temperature (62 °C) and extension temperature (72 °C) and shorter extension time (30 sec.). These reactions were then run on a formamide sequencing gel prepared as the gel above, except using only 8.5 mL H_2O, adding 31.5 mL formamide, and increasing the TEMED to 125 μL. After running,

the gel was rinsed with a 20% methanol, 5% acetic acid solution for 10 min. before drying and autoradiographing as usual.

2. Automated (LI-COR) sequencing:

In spite of several attempts to sequence clones from the 5' region of *SOX1* with automated sequencing, I was not successful. Automated sequencing was used successfully only for the fetal brain library clones I obtained from the *SOX1* 3' region.

a. Mix template DNA and IR-labeled primer (M13 F or M13 R)

5 μL DNA
1 μL primer
11 μL H2O
17 μL total

b. Place 4 μL aliquots of DNA/primer mixture into four tubes, labeled GATC.

c. Add 2 μL of each GATC reagent into labeled tubes. Cover with 1 drop oil.

d. Cycling: 5 min. at 92°C, followed by 35 cycles of 30 sec. at 92°C, 30 sec. at 55°C, 3 min. at 68°C .

e. Termination: Add 3 μL LI-COR stop dye to each tube. Place at –20°C wrapped in foil to prevent photodegradation of product.

f. Submit to automated sequencing facility for analysis on LI-COR sequencers.

O. Rapid amplification of cDNA ends (RACE)

RACE reactions were performed on Marathon-Ready cDNA by Clontech according to manufacturer's instructions. Gene-specific primers used are indicated in **Table 8** in the Results section.

P. Primer extension

1. Primer labeling
    a. For each primer, prepare a labelling reaction as follows:
        1.5 µL H_2O
        2 µL 5x T4 kinase exchange buffer
        1 µL 0.1 M DTT
        1 µL spermidine (1mM)
        1 µL 100 ng/ µL primer DNA
        0.5 µLT4 kinase
        3 µL 10 µCi/ µL γ^{32}P-ATP
        10 µL total
    b. Incubate for 1 hr. at 37°C.
    c. Stop by adding 2 µL 0.5 M EDTA + 50 µL TE buffer and heating for 10 min. at 65°C.
    d. Remove unincorporated nucleotides by passing reaction through a Qiagen spin column according to manufacturer's instructions. Elute in 200 µL DEPC H_2O and count 2 µL for specific activity.

2. Hybridization
    a. Mix 10 µL RNA (10 – 50 µg total RNA), 1.5 µL PE hybridization buffer, and 3.5 µL labelled probe.
    b. Incubate for 90 min. at 65°C. Cool slowly to room temperature.

3. Extension
    a. Prepare the extension mixture by combining the following reagents:
        0.9 µL 1 M Tris-Cl
        1.8 µL Mg Cl2
        0.25 µL DTT
        3.33 µL 2 mM 4x dNTPs
        17.1 µL H2O
        6.75 µL Actinomycin D, 1 mg/mL
        30.33 µL Total
    b. Add extension mixture to hybridization reaction for a total volume of 45.33 µL.
    c. Incubate for 1 hr. at 42°C.

4. Stop

a. Add 105 µL RNase reaction mix to extension mixture for a total volume of 150.33 µL. Incubate 15 min. at 37°C.

b. Add 15 µL of 3 M NaOAc. Transfer to 1.5 mL eppendorf tube containing for a total volume of 150 µL of phenol(25) : chloroform (24) : isoamyl alcohol (1). Shake vigourously, spin one minute.

c. Remove aqueous layer to fresh tube.

d. Add 400 mL ice cold 95% ethanol. Spin 10 min. at 13000 rpm.

e. Remove ethanol. Wash in 100 µL cold 70% ethanol. Dry. Add 5 µL stop dye. Store at 4 – 8 °C until needed.

f. Denature at 70°C. Run reaction on a sequencing gel, adjacent to a sequencing reaction carried out using the same primer.

g. Dry and autoradiograph.

Q. Cell culture

The cell line, NTera2 cl.D1, was ordered from American Type Culture Collection (catalogue number CRL-1973). This cell line is a pluripotent human embryonal carcinoma line originally derived in 1981 from cells isolated from metastatic tumour tissue in a 22 year old male with primary embryonal carcinoma of the testis (Andrews et al. 1984). The cells were cultured according to instructions. Cells were grown at 37°C under 10% CO_2 in Dulbecco's modified Eagle medium with 4 mM L-glutamine adjusted to contain 1.5 g/L sodium bicarbonate, 4.5 g/L glucose, and 1.0 mM sodium pyruvate, 90%; FBS, 10%. Media was changed every 3 – 4 days.

These cells undergo differentiation along a neuronal pathway when treated with retinoic acid (Pleasure et al. 1992). Accordingly, established cultures were reseeded at a density of 1 x 10^6 per 75 cm^2 in medium containing 10^{-5} M all-trans-retinonic acid. Retinoic acid was maintained in the medium for a period of 3 weeks. After this time, cells were replated onto Matrigel diluted 1:60 following the manufacturer's instructions. Mitotic inhibitors were added to the medium, including 1 μM cytosine arabinoside, 10 μM fluorodeoxyuridine, and 10 μM uridine. Unfortunately, cultures abruptly failed at this point and were discontinued.

R. Fluorescence *in situ* hybridization (FISH) on metaphase spreads

This protocol is based on the one used in the University of Alberta cytogenetics research lab which was modified from Fan et al. 1990.

1. Preparation of metaphase chromosome culture
    a. Collect blood from one healthy male volunteer in heparin tubes.
    b. Culture for 72 hours at 37°C in upright flasks in tissue culture incubator as follows:
        6 mL blood
        42 mL RPMI
        10 mL FCS
        1.7 mL fresh PHA
    c. Arrest by adding 0.6 mL colcemid and incubating for 1 hr. at 37°C.
    d. Spin 20 min. at 800 rpm in 50 mL conical tubes. Discard supernatant, resuspend pellet.
    e. Add 30 mL of 0.075M KCl (hypotonic solution) at 37°C. Mix and then incubate 25 min. at 37°C.
    f. Add 2 mL fixative (fresh 3:1 methanol:acetic acid) and spin 10 min. at 1000 rpm.
    g. Remove all but 15 mL of supernatant. Fill tube with fresh fixative and mix. Spin 10 min. at 1000 rpm.
    h. Repeat fixation (step "g") until pellet is white and supernatant clear (about 5 – 8 times).
    i. Store pellet in fixative O/N at 4°C.

2. Preparation of metaphase chromosome slides
    a. Soak microscope slides in 95% ethanol for 30 – 60 min. Place in sterile H_2O and chill at -20 °C (do not freeze).
    b. Drop a few drops of metaphase culture in fixative onto slightly angled cold wet slide from a height of 40 – 50 cm.
    c. Blow on slide gently. Immediately place directly over 65 – 70°C steam for 5 min.
    d. Dry, label, check slides for presence of metaphase chromosomes with microscope, then place in dessicator for one week. For longer term storage, place at -70°C.

3. Labeling probe (carried out on P1 clone #11396)

a. Combine probe DNA and H_2O. Boil 10 min. then place on ice for 5 min.

2 µL DNA (about 2.5 µg)

6.3 µL H_2O

b. Add the following reagents, on ice:

2 µL of each non-labeled nucleotide (A, G, T)

0.5 µL of non-labeled dCTP

2 µL reaction buffer

1 µL Klenow enzyme

2.5 µL of biotin-labelled dCTP

20 µL total

c. Incubate reaction at 37°C for at least 2 hours.

d. Precipitate probe by adding the following and placing at -20°C O/N:

10 µL salmon sperm DNA

15 µL Cot I DNA

4.5 µL NaOAc

112.5 µL 95% ethanol

142 µL total

e. Preparation of probe for hybridization: spin 15 min. at 4°C, rinse in 70% ethanol, air dry, dissolve in 90 µL H_2O. Vortex for 3 hrs.

4. Hybridization

a. RNase treatment: add 200 µL of 2 mL 2xSSC, 20µL RNase A to each slide. Incubate 1 hr. at 37°C in moist chamber. Rinse twice in 2xSSC. Dehydrate with successive rinsing in 70%, 95%, and absolute ethanol. Air dry.

b. Denaturation: soak slides for 3 min. at 70°C in 70% formamide, 2xSSC. Dehydrate with successive rinsing in 70%, 95%, and absolute ethanol at -20°C.. Air dry.

c. Add 200 µL Proteinase K (0.06 µg/ml) per slide. Incubate 8 min. at 37°C in moist chamber. Dehydrate with successive rinsing in 70%, 95%, and absolute ethanol. Air dry.

d. Add 1 mL hybridization solution per probe. Vortex thoroughly. Denature probe for 5 min. at 70°C, then ice 2 min. Add 100 µL probe per slide. Incubate O/N at 37°C in moist chamber.

5. Washing: place slides in holders and perform washes with shaking in a 45°C waterbath as follows.
    a. 2 x 15 min. in 50% formamide, 2xSSC
    b. 10 min. in 2xSSC
    c. 10 min. in 1.5xSSC

6. Immunofluorescence: slides are subjected to a series of washings in buffer and treatments with fluorescent reagents applied directly to each slide as follows.
    a. Soak in BN buffer, 10 min. at room temp.
    b. Add 100 µL Pre-FAD, 5 min. at room temp.
    c. Add 100 µL FAD, 30 min. at 37°C.
    d. Wash in BN buffer, 10 min. at 45°C.
    e. Add 100 µL Pre-BAAD, 5 min. at room temp.
    f. Add 100 µL BAAD, 30 min. at 37°C.
    g. Wash in BN buffer, 10 min. at 45°C.
    h. Add 100 µL Pre-FAD, 5 min. at room temp.
    i. Add 100 µL FAD, 30 min. at 37°C.
    j. Wash in BN buffer, 10 min. at 45°C.
    k. Add 45 µL DAPI/PI. Dry 5 min.
    l. View under fluorescent microscopy.

S. Radiation hybrid (RH) mapping

1. *SOX1* primers used (200 ng each): 1611, 2077, expected product size 482 bp.
2. Templates used (25 ng each):
    93 radiation hybrid panel samples (Research Genetics, Inc.)
    HFL (human fetal lung)
    A23 (hamster)
    H_2O
3. PCR conditions:
    60°C annealing temperature
    30s:30s: 60s denature:anneal: extension times
    7.5% DMSO
    25 µL reaction volumes
4. Gel conditions: 1.5% agarose TBE gel run for 2 hrs. at 100V
5. Each lane was scored as positive, negative, or ambiguous, and analysis of these results was performed using Whitehead Institute for Biomedical Research server.

T. Table 5. World Wide Web Resources

SITE DESCRIPTION	URL Address
BLAST (Basic Local Alignment Search Tool) is a set of similarity search programs designed to explore all of the available sequence databases regardless of whether the query is protein or DNA.	http://www.ncbi.nlm.nih.gov/BLAST
OMIM (Online Mendelian Inheritance in Man) is a database of human genes and genetic disorders complied by Dr. Victor A. McKusick and colleagues.	http://www.ncbi.nlm.nih.gov/Omim
CBS (Centre for Biological Sequence) analysis is a jumping off site for various nucleotide and amino acid analysis programs compiled by the Technical University of Denmark, including Promoter 2.0 and HMMGene.	http://www.cbs.dtu.dk/services/
PROSITE is a database of protein families and domains. It consists of biologically significant sites, patterns and profiles that help to identify to which known protein family (if any) a new sequence belongs.	http://expasy.cbr.nrc.ca/prosite/index.html
TIGR The Human Gene Index (HGI) integrates research results from international human gene research projects.	http://www.tigr.org/tdb/hgi/hgi.html
PSORT predicts protein localization sites in cells by applying the stored rules for various sequence features of known protein sorting signals.	http://psort.nibb.ac.jp/
PROFILE SCAN uses the pfscan program to search a protein sequence against available databases (PROSITE)	http://www.isrec.isb-sib.ch/software/PFSCAN_form.html

U. Patient descriptions

Blood samples were collected from patients and members of their families according to standard methods. All participating individuals gave full and informed consent according to ethical guidelines followed by the referring physicians and approved by the University of Alberta Ethics Review Board. Patient phenotypes are summarized in **Table 6**.

More clinical information was available regarding the campomelic dysplasia patient than other patients. This female patient was the child of a non-consanguineous Caucasian couple (father age 32 years, mother age 28 years). Routine ultrasound study at 19 weeks of gestation showed no abnormalities, but at 34 weeks the fetus was small for gestational age. Detailed ultrasound study at this time demonstrated abnormalities of the cervical vertebrae, short limbs, and a large head. Labour was induced at 39 weeks gestation and delivery was by Caesarian section in response to fetal distress.

The birth weight was 3,155 g (50th centile), the head circumference (OFC) 36 cm (90th centile), and the length was 46.5 cm (25th centile). Apgar scores were 6 at one minute and 9 at 5 minutes. The infant had a large head with frontal bossing and a prominent capillary hemangioma over the forehead and both upper eyelids. Her facial profile was flattened, and she had a broad nasal bridge and an upturned nose. She had short broad fingers and single palmar creases bilaterally. Her elbows did not extend fully, and there was acromelic shortening of both upper and lower limbs. Below the knee, there was a prominent angulation of the tibia with an overlying skin dimple, and there was talipes equinovarus at the right ankle. Her heart was normal, but she had a small chest. She had marked laryngomalacia with paradoxical respiration when crying. Her neuromuscular tone was reduced and there was generalized ligamentous laxity.

A CT scan of the head found skull asymmetry, and radiological examination showed an enlarged cranium. The cervical spine showed fusion of the bodies of C3 and C4, and C4 and C5 were hypoplastic, wedge-shaped, and exhibited subluxation. The odontoid process was oddly shaped, and the vertebral bodies were foreshortened along the anterior-posterior axis. The patient had hypoplastic scapulae and thoracic kyphosis. The left hip was dislocated and a pseudoacetabulum had formed on the left. The left proximal femoral and tibial epiphyses were absent. The ulna was foreshortened relative to the radius, giving

rise to a 'main en bayonet' deformity. There was delayed ossification of the small bones of the hand, and the capitate and hamate bones were fused.

Early infancy was complicated by life-threatening respiratory obstructive episodes, which gradually resolved with age. At age one year, she was sitting with support and reaching out to grasp objects. She recognized her relatives. The peripheral reflexes were decreased in the lower limbs. Ophthalmologic examination showed hypermetropia. At age two, MRI showed non-progressive hydrocephalus. The 'main en bayonet' appearance became more pronounced as she grew. Subluxation developed at the left knee. Her height and weight dropped to four standard deviations below the mean for age, in contrast to her head circumference at two standard deviations above the mean. At the age of four, the child began to have episodes of inconsolable crying, apparently in pain. Imaging of the complete spine and extensive investigations could not uncover a cause. After discussion with the parents, a decision was made to treat palliatively. The child died at age 5 years.

Table 6 Legend:

*‡§ Members of the same family are identified with the same mark.

\+ indicates positive diagnosis for indicated condition, blank indicates no information for condition

Abbreviations: ID identification code

MO microphthalmia
MC microcornea
CA cataract (unspecified type)
ZC zonular cataract
G glaucoma
AN aniridia
ARS Axenfeld – Rieger syndrome
ARA Axenfeld Rieger anomaly
IGD iridogoniodysgenesis
PT Peter anomaly
CED corneal endothelial dystrophy
ASD anterior segment mesenchymal dysgenesis
UA unaffected
UK unknown

Table 6. Phenotypic information for individuals tested for *SOX1* mutations.

ID	MO	MC	CA	ZC	G	AN	ARS	AR	IGD	PT	CED	ASD	UA	UK
919														+
920														+
922*			+											
923*			+											
921‡		+	+		+									
924‡													+	
925‡													+	
926			+		+									
927				+										
928	+		+		+									
929														+
930													+	
1137	+													
1138	+													
1139	+													
1140														+
1141														+
SD§			+											
MD§													+	
LD§	+		+											
AP	+		+											
HF	+		+											
HM	+		+				+							
MA							+							
CA						+								
SB							+							
JC							+							
GC								+						
MF							+							
MF2									+					
AG							+							
EG					+									
BH											+			
RH									+					
JJ												+		
JJ2							+							
GL							+							
SM											+			
NN									+					
MP							+							
LS							+							
JS										+				
CS									+					
MW										+				

RESULTS

A. Cloning *SOX1* and *SOX2*

1. Library screening

Information about murine *Sox1* and *Sox2* expression in the developing neural tissues of embryonic mice suggested that the transcripts of these genes might be present in human fetal brain (R. Lovell-Badge, personal communication, 1994). Accordingly, a 450 bp murine *Sox2* probe (a generous gift of P.N. Goodfellow, Cambridge University) was chosen to screen a human fetal brain cDNA library to obtain putative human *SOX1* and *SOX2* cDNA clones. Since Sox2 is over 50% identical to Sox1 at the amino acid level, it was hoped that a low stringency screen using the *Sox2* probe might also recover a *SOX1* cDNA. A summer student (O. Tchernaia), initiated the primary screening of approximately 2 x 10^6 plaques from a Stratagene human fetal brain cDNA library.

I selected ten positive plaques from the primary screening of the cDNA library, designated A through J, for further characterization. After secondary screening with the same murine *Sox2* probe used in the initial screening, seven positives were carried forward for tertiary screening. Five positives resulted from the tertiary screening, namely, A, B, C, E, and H. Secondary lifts from the D, F, G, I, and J plaques were re-tested with the *Sox2* probe; G, I, and J were confirmed to be negative, but D and F resulted in weak positives. All five of the initial positives, as well as the two weak positives (D and F), were subjected to conversion into a plasmid form using the Stratagene excision protocol for the λ-ZAP phagemid. The excision protocol failed to produce colonies for strains D and H, resulting in five strains in plasmid form for further analysis, A, B, C, E, and F. **Table 7**

summarizes the results of the human fetal brain cDNA library screening with the murine *Sox2* probe.

Table 7: Human fetal brain cDNA library screening with murine *Sox2* probe

Plaque	Primary	Secondary	Tertiary	Excision
A	+	+	+	+
B	+	+	+	+
C	+	+	+	+
D	+	-/+	N/D	-
E	+	+	+	+
F	+	-/+	N/D	+
G	+	-	N/D	N/D
H	+	+	+	-
I	+	+	-	N/D
J	+	+	-	N/D

Legend: N/D = not done; -/+ = weak positive

2. Clone restriction mapping and subcloning

As an initial step in characterizing the putative *SOX* clones, restriction mapping was performed using a variety of enzymes. Insert sizes were determined from *Eco* RI digests. Comparison of the band patterns with frequent cutting enzymes suggested that clones B and E may be overlapping fragments of the same cDNA. **Figure 8** depicts composite restriction maps for some of the enzymes used, and indicates the total insert sizes for the five clones.

Southern blots were made from gels of digests of the five clones, and these were probed with the murine *Sox2* probe and an *SRY*-box probe. Two of the clones, B and E, showed very strong hybridization to both probes, suggesting they might be *SOX1* or *SOX2* clones. The remaining three clones, A, C, and F, hybridized only

weakly. Clones A, B, E, and F were subdivided using the enzymes *Eco* RI and *Pst* I into the subclones indicated on **Figure 10** (a summer student, S. Hepfner, subcloned A and F). Each of the subclones was then used for sequence analysis.

3. Clone sequencing

Manual sequencing was carried out on the subclones made from human fetal brain cDNA library clones A, B, E, and F, and the DNA sequences were analyzed using the Basic Local Alignment Search Tool (BLAST) on-line program (URL: http://www.ncbi.nlm.nih.gov/blast). End-sequence from subclones B5 and E3 matched human *SOX2* sequence, just downstream of the HMG box, indicating that these clones were most likely derived from *SOX2* cDNAs (indicated on **Figure 10**). These results confirmed that clones B and E did overlap, and shared at least a common region around the *Pst* I site present in *SOX2*. Clone B was found to be chimeric, since sequence from the opposite end of subclone B5 gave an excellent match to the human *Neuroendocrine specific protein* (*NSP*) gene, located on chromosome 14 (*SOX2* is located on chromosome 3) (Roebroek1993). Clone E may or may not be chimeric, since sequences from the end of subclone E3 opposite the *SOX2* end did not match any sequences in the database. Unfortunately, the genomic region around *SOX2* has not yet been sequenced, and the available *SOX2* sequence is limited to little more than the coding region of the gene, making it very difficult to tell whether other sequences are in the same region or even on the same chromosome.

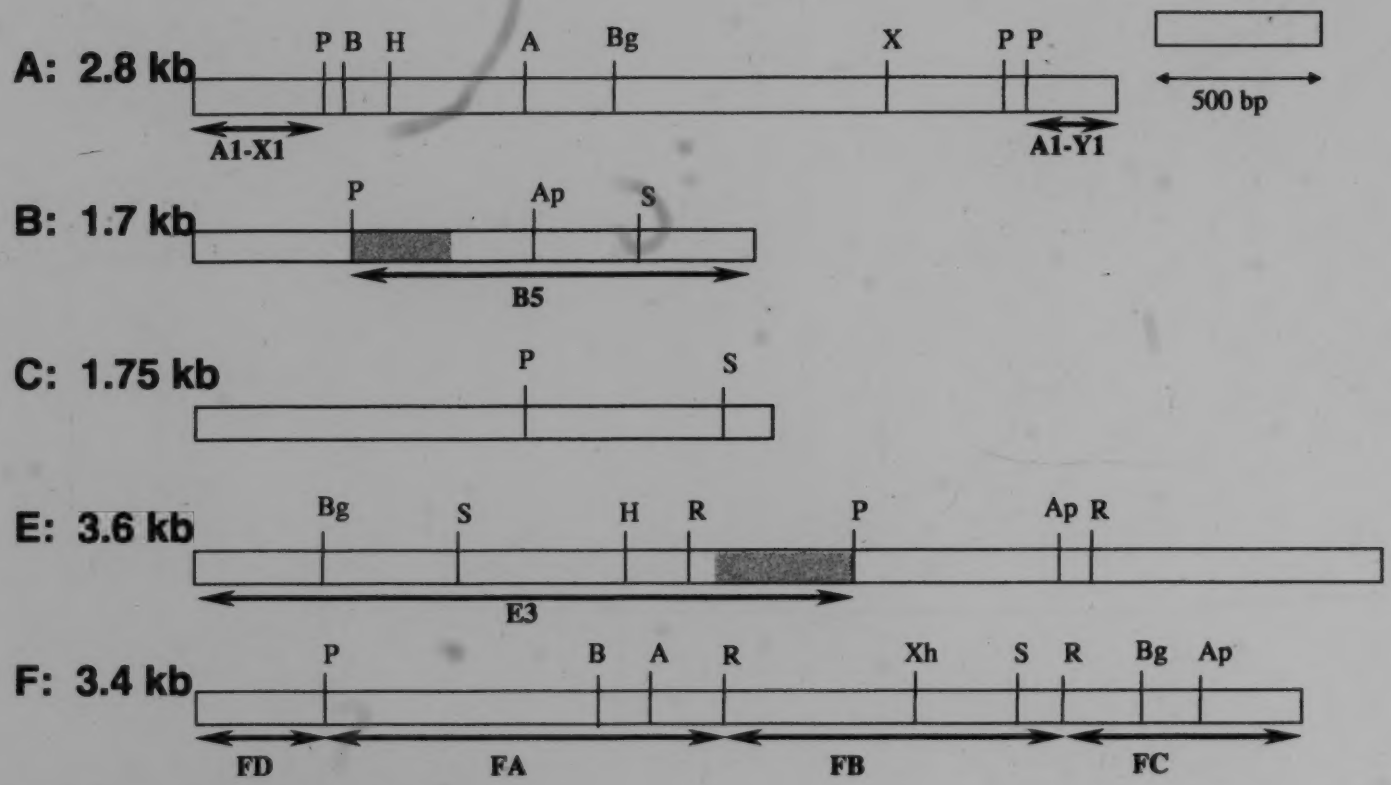


Figure 10: Restriction Maps of SOX Clones

Legend: Five SOX clones were mapped with *Acc* I (A), *Apa* I (Ap), *Bam* HI (B), *Bgl* II (Bg), *Eco* RI (R), *Hind* III (H), *Pst* I (P), *Sca* I(S), *Xba* I (X), *Xho1* (Xh), and other enzymes (not indicated). Arrows indicate subclones used for sequence analysis. Sequenced regions matching *SOX2* sequence by BLAST are shaded.

Sequences from clones A and F did not initially match any sequences in the database. Later analysis tested end-sequences against the high throughput genomic database, which allowed tentative elucidation of the full sequence for both A and F. Clone A matches sequence from chromosome 20, sequencing clone RP4-591C20, nucleotides 76334 - 79168. Although human *SOX22* has been mapped to chromosome 20p13, its sequence does not match the same genomic contig as clone A. Also, no *SOX*-like sequences were found within clone A, suggesting that it does not correspond to *SOX22*. Clone A does give a strong BLAST match to a breast carcinoma-associated D52 protein family gene called hD54 (accession number AF004430).

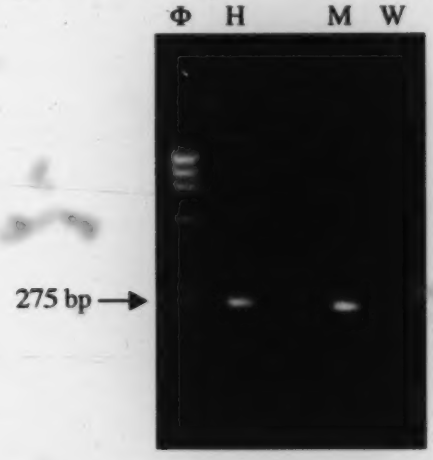
Clone F matches sequence from chromosome 4, sequencing clone RP11-77403, nucleotides 30278 – 32761, 98774 – 99384, 100702 – 100881. The breaks in the genomic sequence may correspond to intronic sequence missing from the cDNA clone. A small amount of sequence from clone F, about 70 bp at the very end of subclone FC, matches sequence from chromosome 2. Clone F may be chimeric, including a tiny region from chromosome 2 and the majority from chromosome 4. To date, no human *SOX* genes have been mapped to chromosome 4, suggesting that clone F is not a known *SOX* gene. Since no *SOX*-like sequences were detected in clone F, it probably represents another, non-*SOX* family, gene. Clone F also contains a repetitive element, but even without this part of the sequence, it does not yield any significant hits by BLAST searching the non-redundant database.

B. PCR-based strategies to isolate *SOX1*

1. PCR amplification of *SOX1*

Because none of the fetal brain clones represented human *SOX1*, a PCR approach was undertaken to try to amplify *SOX1* from human genomic DNA. PCR-based cloning methods are most effective for genes which lack introns, like *SOX1*, since the coding sequence is contiguous with the genomic sequence. I started with three primers designed by M.Walter (A,B, C) for the 3' region of the murine *Sox1* gene, as well as a degenerate HMG box primer pair (refer to **Appendix C** for primer sequences). After unsuccessful attempts to amplify human *SOX1* this way, I then used the murine *Sox1* sequence to design primers (D, E, F, and G) to amplify the human *SOX1* HMG box region. The primer set F/G amplified a band of the correct size, 275 bp, from both human and mouse DNA (**Figure 11**). These bands were excised from the gel, sequenced, and confirmed to comprise the HMG box regions of human and mouse *SOX1*.

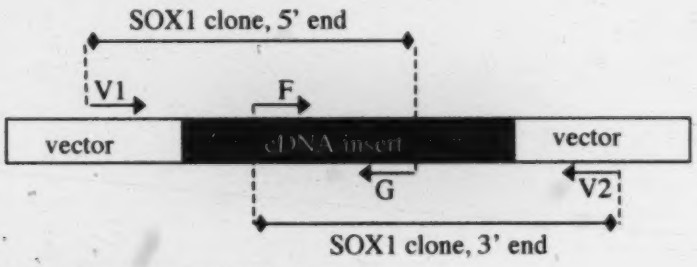
Figure 11. PCR amplification of F/G *SOX1* fragment



Legend: Φ, ΦX174/Hae III marker H, human genomic DNA
W, water M, mouse genomic DNA

An attempt was made to take advantage of the F/G primer pair to obtain larger 5' and 3' *SOX1* clones from a cDNA library by using overlapping PCR reactions. This strategy is based on RACE methods of obtaining cDNA clones for 5' or 3' ends. Each of the two gene-specific primers (F and G) is paired with a vector-specific primer (V1 and V2) to amplify an intervening region of the cDNA. Positives are confirmed by testing with the internal gene-specific primer pair. A diagram of this method is shown in **Figure 12**.

Figure 12. Library RACE method for cDNA cloning



Legend: V1, V2 vector-specific primers
F, G SOX1-specific primers

The library used for this screen was a human retinal library in the λgt11 vector, chosen because a lens library was not available. Although this method produced several products, which were cloned using a TA cloning kit, none represented *SOX1* sequence. A likely explanation for this failure is the lack of *SOX1* expression in the human retina. *In situ* hybridization using a *SOX1* probe subsequently confirmed the lack of expression of this gene in the retina (see Introduction **Figure 6**, Kamachi et al. 1998).

2. P1 Library Screening for *SOX1*

The F/G primer set was used to procure a *SOX1* genomic clone by commercial PCR screening. Large insert genomic libraries in bacterial artificial chromosomes (BACs) or bacteriophage P1 vectors (P1s) were chosen for their noted stability in comparison to yeast artificial chromosomes (YACs), and because of the chance of obtaining the full *SOX1* sequence along with substantial flanking domains from a single clone. A BAC library screen by Research Genetics, Inc., did not yield a clone. A subsequent P1 library screen by Genome Systems, Inc., produced two P1 isolates, numbers 11396 and 11397, which were positive for the *SOX1* F/G primers. This positive result was confirmed by PCR on DNA from the P1 clones.

Once these P1 clones were received, additional PCR-based cloning attempts were made using P1 DNA as a template and the previously developed primers (A through G) as well as newly designed primers (SRY-5' and PA-3') to obtain additional *SOX1* subfragments for sequencing. In comparison to a genomic DNA template, the P1 DNA offered a much more specific starting material, theoretically allowing amplification to proceed more readily. Several products of approximately appropriate sizes were cloned using the TA cloning kit. Unfortunately, these proved to be PCR artifacts, as sequencing revealed that no *SOX1*-containing products were amplified.

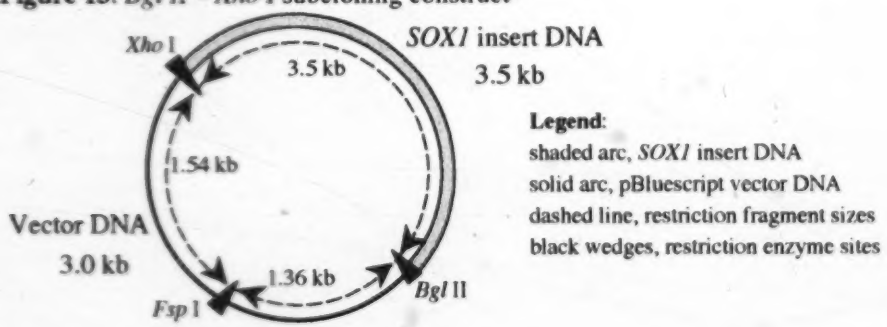
3. P1 subcloning and sequencing

In order to determine the sequence of the *SOX1* gene, P1 number 11396 was subcloned to obtain suitably sized fragments for DNA sequencing. Numerous digests, gels and Southern blots were made in order to isolate a P1 fragment containing the *SOX1* gene that could be subcloned and sequenced. Based on

restriction digest analysis, the approximate size for the entire P1 is 70 - 80 kb, of which 16 kb is vector. A 3.5 kb *Bgl* II – *Xho* I fragment which hybridized strongly to a probe from the *SOX1* HMG box was identified, and this fragment was cloned into the pBluescript KS II(-) vector for sequencing and further subcloning. This initial cloning was initiated by a summer student (C. Rudnisky).

I then adopted a shotgun strategy for subcloning the entire *Bgl* II – *Xho* I fragment into smaller pieces. Insert DNA was digested with *Sau3A I* and cloned into the pBluescript KS II(-) vector at its *Bam* HI site. After finding an unexpectedly high number of vector subclones, an additional step was added during the purification of the *Bgl* II – *Xho* I insert DNA to avoid vector contamination. Since the *Bgl* II – *Xho* I fragment was 3.5 kb and the vector size was 2.9 kb, these bands migrated close together on electrophoresis, and the insert band was contaminated with vector DNA. To avoid such close migration, the plasmid was cut with *Fsp* I, an enzyme which cut the vector but not the insert. The insert DNA could then be more easily purified from a gel without vector contamination. A simple diagram of the *Bgl* II – *Xho* I construct used for subcloning is shown in **Figure 13**.

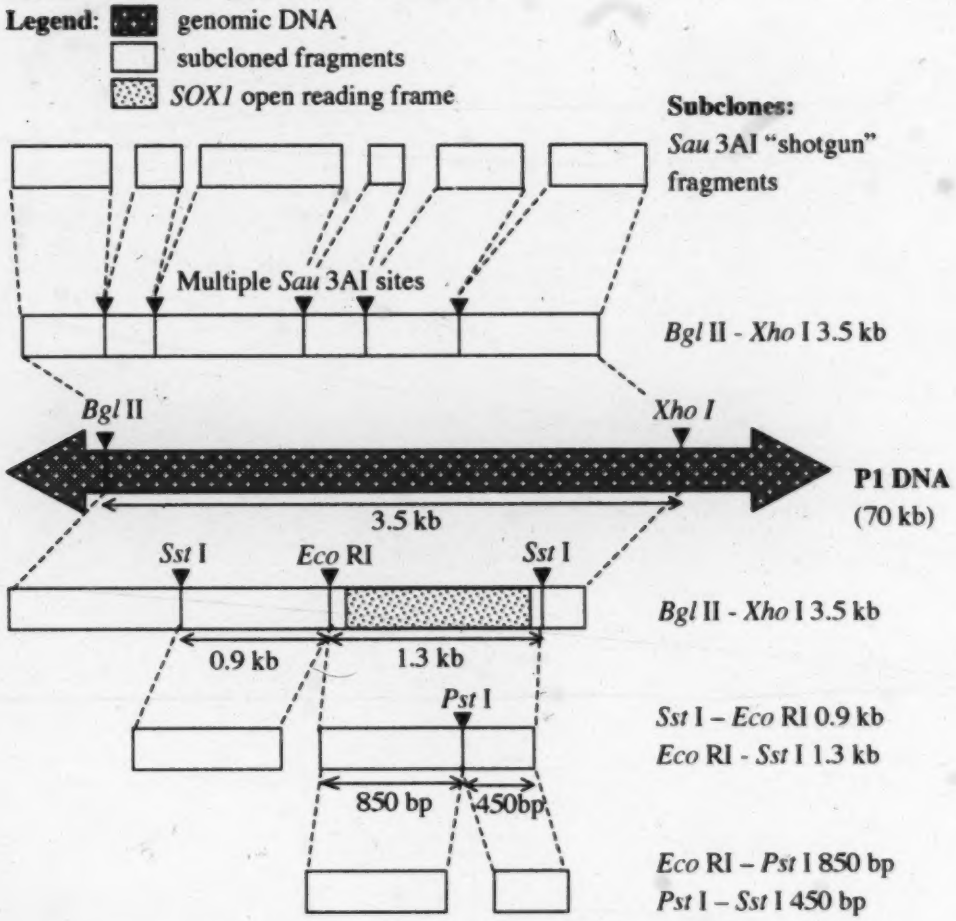
Figure 13. *Bgl II – Xho I* subcloning construct



The shotgun subcloning strategy effectively allowed sequencing of a large portion of the3.5 kb *Bgl* II – *Xho* I *SOX1* clone. To fill remaining gaps in the sequence of the fragment, new primers were designed based on the newly determined human *SOX1* sequence, both to use for PCR amplification from the P1 DNA and to use as sequencing primers. These new primers included HP-12, 355, 473, 795, 1611, 1632, 1889, 2077, 2149, and others, and were used in a variety of combinations for PCR. Although some progress was made with these new primers, sequencing from P1 DNA was not easy, and better results inevitably came from sequencing from universal primers into subcloned plasmid DNA. For this reason, another series of experiments was undertaken to produce additional subclones of the *Bgl* II – *Xho* I 3.5 kb *SOX1* clone to use for direct sequencing.

The *Bgl* II – *Xho* I 3.5 kb *SOX1* clone was subdivided into *Eco* RI – *Sst* I fragments, and two subclones, approximately 1300 and 900 bp in size, were recovered. The 1300 bp *Eco* RI – *Sst* I clone, which contained within it the entire *SOX1* open reading frame, was then further divided into two subclones of 850 and 450 bp, using the enzyme *Pst* I. A diagram of the entire subcloning process is given in **Figure 14.**

Figure 14. Subcloning *SOX1* from P1 DNA

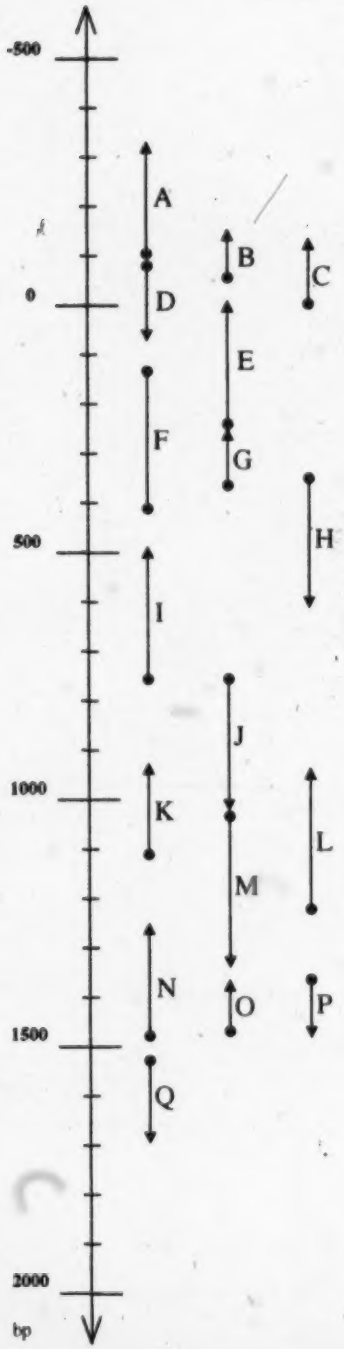


Sequencing of the *SOX1* region was fraught with difficulty. Much of this trouble can be attributed to the very high GC content of the *SOX1* gene, about 76% over the coding region. In addition to producing a large number of different overlapping fragments for sequencing templates, a number of technical strategies were used to successfully sequence the entire region. These alterations to standard sequencing conditions included: using d-ITP instead of d-GTP, adding 5% DMSO, quadrupling the amount of enzyme used, adding a 5 minute denaturing step at the beginning of the reaction, doubling the number of reaction

cycles, decreasing the amount of primer, and running the reactions on formamide-containing gels. Even under these stringent conditions, sequencing gels often required long exposure times on autoradiography in order to detect results.

A contig of *SOX1* sequence, depicted in **Figure 15**, was assembled from the various sequences determined from subclones via universal and internal primers. The contig represents approximately 2.2 kb of the human *SOX1* genomic DNA, including a 1173 bp putative coding region and about 1000 bp of flanking region.

The human *SOX1* DNA sequence with its putative translation is depicted in **Figure 16**. Computer analysis of the sequence was carried out for gene features (URL: http://www.hgsc.bcm.tmc.edu/SearchLauncher). No intronic sequence has been identified, consistent with *SOX1* being a single exon gene. No transcription start signal was found in the 5' sequence, and neither was a polyadenylation signal found in the 3' sequence. The absence of these transcription initiation and termination signals implies that the untranslated region of *SOX1* extends beyond the 2.2 kb of sequence determined from this contig, likely in both directions.



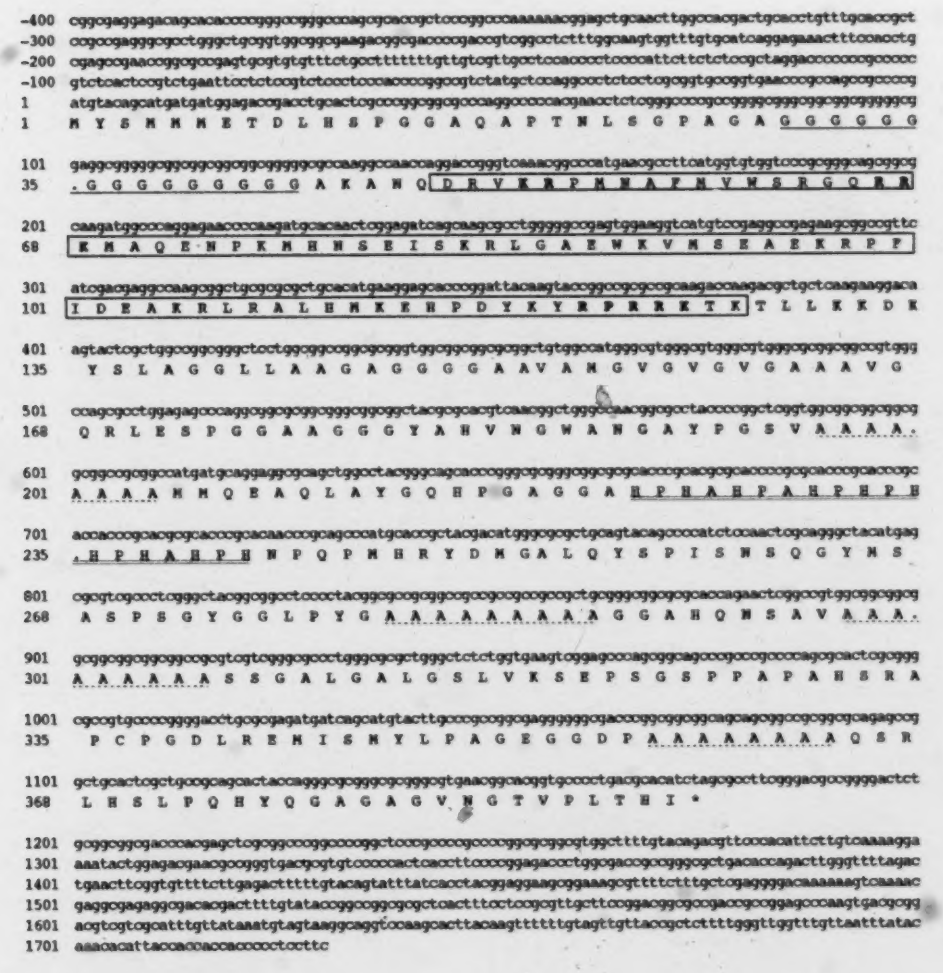
	TEMPLATE	PRIMER
A	*Sau3A I* #10	UCP
B	P1	473
C	P1	473
D	*EcoR I – Pst I* 850 bp	R
E	*Sau3A I* #7	F
F	F/G PCR product	UCR, RCP
G	*Sau3A I* #8	R
H	*EcoR I – Pst I* 850 bp	795
I	*EcoR I – Pst I* 850 bp	F
J	*Pst I – Sst I* 450 bp	R, KS
K	P1	1632
L	*Pst I – Sst I* 450 bp	F
M	*Sau3A I* #11	F
N	*Sau3A I* #11	R
O	*Sau3A I* #2	RCP
P	*Sau3A I* #2	UCP
Q	P1	1889

Figure 15. *SOX1* sequencing contig
Legend: Scale position 1, first A of start codon; position 1174, first T of stop codon. Arrows indicate direction of read.

The *SOX1* sequence (**Figure 16**) encodes a predicted polypeptide of 391 amino acids (estimated MW≈39 kDA). Of three possible initiation codons, the sequence flanking the first predicted start codon, GCC GCC CCG ATG T, correlates well with the Kozak consensus sequence, (GCC) GCC PuCC ATG G. However, at position -3, the *SOX1* sequence has a pyrimidine (C) instead of a purine, which is typically conserved in vertebrate translation initiation sequences (Kozak, 1987). SOX1 exhibits several significant protein motifs. As in other *SOX* genes, the 79 amino acid HMG box (aa 49 – aa 127) contains within it a bipartite nuclear localization signal (aa 52 – 53 and aa 66 – 68) and a basic cluster nuclear localization signal (aa 121 – 125) (Poulat et al. 1995). As well, *SOX1* is the only known *SOX* gene to contain a histidine-proline rich paired-type repeat (aa 223 – aa 241). *SOX1* also contains a polyglycine stretch (aa 29 – aa 43) preceding the HMG box, and four polyalanine tracts (aa 197 – aa 204, aa 280 – aa 288, aa 298 – aa 306, aa 357 – aa 364) distal to the box. Possible functions of these motifs will be addressed in the discussion in the section, "The *SOX1* gene."

Analysis of the SOX1 predicted protein sequence using PROSITE and CBS for potential sites of modification yielded several possibilities (refer to Materials and Methods for web site addresses). Consensus sequences for phosphorylation were found for serine residues 81, 92, and 368 and for threonine residues 43 and 125. Serine residues 269, 321, 323, and 330 were potential O-glycosylation sites, and asparagine residue 381 was a possible N-glycosylation site. One tyrosine kinase phosphorylation site was found at tyrosines 117 and 119, and numerous myristoylation and amidation sites were predicted. The potential biological relevance of these predictions is completely unknown.

Figure 16. *SOX1* DNA Sequence with Translation



Legend: Numbering of the DNA sequence and the protein translation uses the first A of the starting methionine as position number 1. The HMG domain is boxed, the nuclear localization signals are in boldface, the paired-type repeat is underlined with a double line, the polyglycine tract is underlined with a solid line, and polyalanine stretches are underlined with a broken line.

The DNA and predicted protein sequences determined for human *SOX1* are 95% and 99% identical respectively to the overall murine *Sox1* DNA and predicted protein sequences (refer to comparative alignment **Figures 35, 38,** Discussion). An alternate human *SOX1* sequence proposed by Malas et al. (1997) (EMBL Accession No. Y13436) was only 92% and 96% identical to the murine DNA and protein sequences, respectively. Although discrepancies arising from polymorphic variation, subcloning, or sequencing errors in this highly GC-rich gene might explain the differences between my *SOX1* sequence and that reported by Malas, a recently released human sequencing contig (RP11-310D8, Aug. 8, 2000) supports my sequence at every one of the six points of divergence. I have screened a number of patients' DNA sequences for alterations in the *SOX1* coding region and have not found sequences which correspond to that proposed by Malas. In addition, I have double checked the areas of discrepancy for sequence accuracy with two or more sequencing reactions. Furthermore, the sequence I propose is much closer to the murine *Sox1* sequence. For these reasons, I believe that my *SOX1* sequence is likely more accurate than the one published by Malas and coworkers (1997).

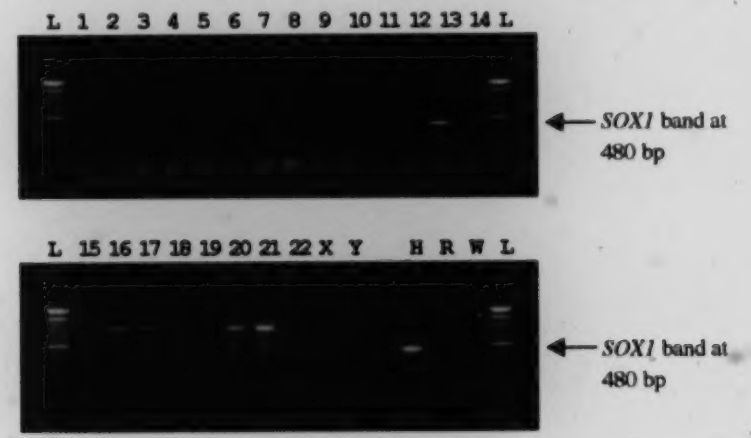
C. Localizing *SOX1*

1. Monochromosomal hybrid mapping

In order to determine the chromosomal location of the human *SOX1* gene, a number of mapping strategies were used, including Alu PCR, monochromosomal hybrid panel PCR, FISH, and RH mapping. A summer student (C. Rudnisky) attempted chromosomal assignment by using Alu PCR to amplify a portion of the *SOX1*-specific P1 clone. The product was used to probe a monochromosomal panel Southern blot, but no chromosome-specific signal was detected.

I then began testing *SOX1*-specific PCR primers on a monochromosomal human-rodent hybrid panel of DNA aliquots, provided by H. McDermid (Dept. of Biological Sciences, University of Alberta). The primer pair F/G, which was used for the P1 library screening, amplified the same-sized band from human and rodent DNA, and so was unsuitable for use on the hybrid panel. I attempted to take advantage of small sequence differences between the human and rodent *SOX1* HMG boxes by cutting the F/G PCR product with a species-specific enzyme. I did not get reproducible results in this way, possibly due to polymorphic variants or strain-specific variants in the rodent DNA. After much testing, I devised primer pair, 1611/2077, which selectively amplified a band of approximately 480-bp from the chromosome 13 lane (**Figure 17**).

Figure 17. Monochromosomal Hybrid Panel PCR for *SOX1* Primers 1611/2077

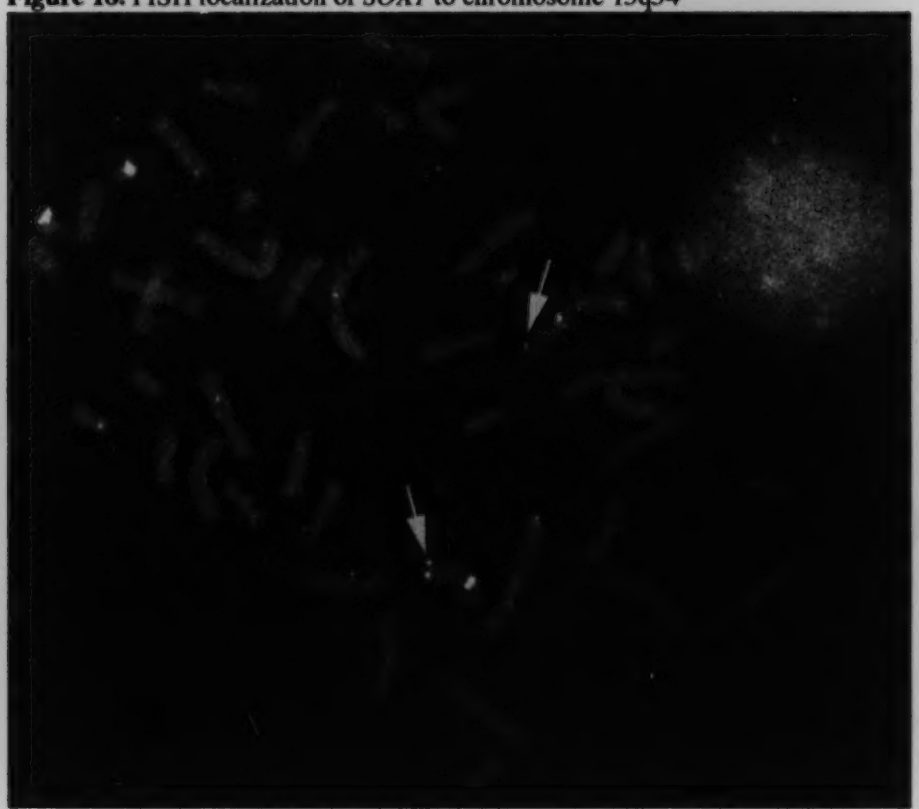


Legend: Chromosome lanes are numbered accordingly
L, 100bp ladder H, human R, hamster W, water

2. Fluorescence *in situ* hybridization (FISH) mapping

After unsuccessful probings with smaller subclones, I biotin-labeled the entire P1 phagemid as a probe for FISH with human metaphase chromosomes. The probe hybridized to the distal end of chromosome 13, indicating that *SOX1* localized to chromosome 13q34 (**Figure 18**). Signals were detected on one or both chromatids of 13q in all metaphase spreads in which signal was present. No additional signal was found elsewhere on 13 or on any other chromosomes. Simultaneous probing with an α-satellite marker confirmed the identity of chromosome 13.

Figure 18. FISH localization of *SOX1* to chromosome 13q34



Legend: Arrows indicate *SOX1* position on chromosome 13q34. Signals from the α-satellite markers are also visible on both chromosomes 13 near the centromere, and on both chromosomes 21.

NOTE TO USERS

Page(s) not included in the original manuscript and are unavailable from the author or university. The manuscript was microfilmed as received.

94

This reproduction is the best copy available.

UMI

D. Expression Studies

1. *SOX1*

To investigate the expression pattern of *SOX1* in human tissues, several approaches were used, including PCR of human cDNA libraries, Northern blot analyses, and RT-PCR reactions from various RNA sources. For comparison, the only study of human *SOX1* expression, to date, reports a moderately abundant 3.9 kb transcript in 16 week embryonic brain mRNA and no expression in the NTera2 cell line, the only two sources tested (Malas 1997). In mouse and chick, where *Sox1* has been studied much more extensively, expression has been seen in the developing central nervous system and in the embryonic eye, specifically limited to the nascent lens (Collignon 1996, Kamachi 1998).

Several human cDNA libraries were tested by PCR for *SOX1* expression using the primer pairs 1889/2077 and 1611/2077. The libraries assayed included craniofacial, lymphoblast, fibroblast, chorioretinal, retinal pigmented epithelium, kidney, iris, and more recently, fetal brain. The fetal brain library persistently showed *SOX1* expression, although the band was often quite weak. Unfortunately, except for fetal brain, the other libraries inconsistently produced only extremely weak bands. Four commercial RACE-ready human cDNA sources were also tested for *SOX1* expression with primers 1611/2077, including placental, Caco (colon cancer), heart, and fetal brain. Weak bands were detected in heart and fetal brain, but subsequent RACE attempts did not produce *SOX1* clones from either of these sources. These inconclusive results may have been due either to poor quality template or to very low *SOX1* expression levels in the tissues tested.

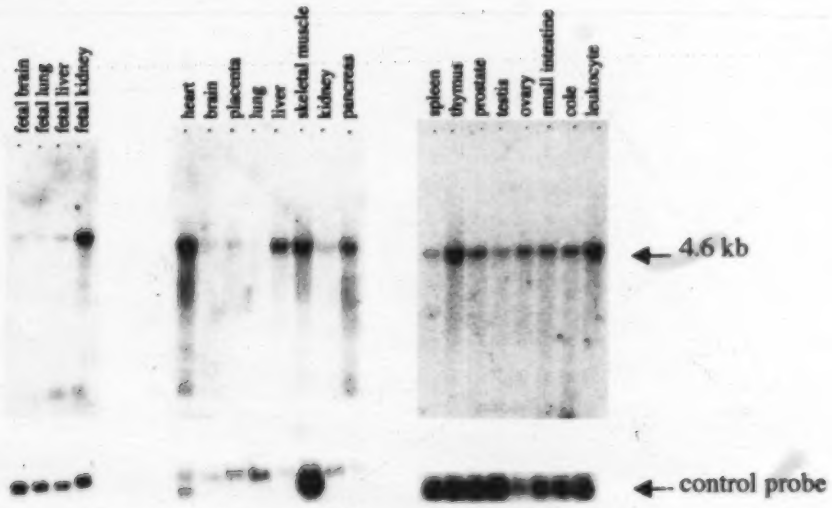
SOX1 expression was also assayed by RT-PCR on RNA from several different sources. RNA from white blood cells, both from fresh blood and from lymphoblastoid cell lines, failed to show evidence of *SOX1* expression by PCR with primers 1611/2077and 12/221. RT-PCR on RNA isolated from fresh human lenses which had been excised for cataract surgery also failed to detect *SOX1* products with primers 1889/2077 or 1611/2077, although a non-specific smear was seen in the RT positive lanes for primers 1889/2077. Finally, RNA isolated from undifferentiated NTera2 cells did show *SOX1* expression by RT-PCR with primers 367F/R and by Northern blot analysis at 4.6 kb. RNA from retinoic-acid treated NTera2 cells did not show *SOX1* expression with the same primer pair, but this RNA was not tested with a positive control and may have been too dilute for analysis. The expression found in NTera2 cells contradicts the results reported by Malas et al. (1997).

Because of the difficulty of finding a *SOX1*-containing RNA source, several attempts were made to detect *Sox1* expression in mouse tissues. I isolated RNA from mouse eye, ovary, testis, and seminal vesicle, and added previously frozen RNA from heart, muscle, kidney and liver to make a mouse multiple tissue Northern blot. The human *SOX1* 450 bp *Pst I – Sst I* fragment isolated by subcloning from the genomic P1 DNA (see **Figure 14**) was used to probe the mouse blot. Significant bands were detected in the eye and kidney lanes, providing evidence of eye-specific expression and underlining the finding of *SOX1* expression in human fetal kidney (discussed below).

Human fetal and adult multiple tissue Northern blots (Clontech) were also probed with the 450 bp *Pst* I – *Sst* I fragment of *SOX1*. This fragment represents the

distal 450 bp of the *SOX1* coding region, and was chosen so as not to include either the HMG box or the PRD-type repeat to avoid cross-detection of related SOX or PRD-type repeat containing genes. Northern blot hybridization detected a 4.6 kb band in all adult tissues examined, suggesting that *SOX1* is ubiquitously expressed in adult tissues. The strongest signal appeared in adult heart, liver, pancreas, thymus, and leukocytes. The same sized band was detected in fetal brain, lung, liver, and most strongly in fetal kidney (**Figure 19**).

Figure 19. *SOX1* expression profile on Northerns with 450 bp *Pst* I – *Sst* I probe



The results of these *SOX1* expression studies are not easy to interpret, since apparent widespread expression by Northern analysis was not found by other means of detection. Because of the high GC content of the *SOX1* gene, which tends to produce secondary structures in the DNA template, PCR reactions proceed with difficulty. Low template levels present in RNA sources in combination with the presence of secondary structures may make RT-PCR

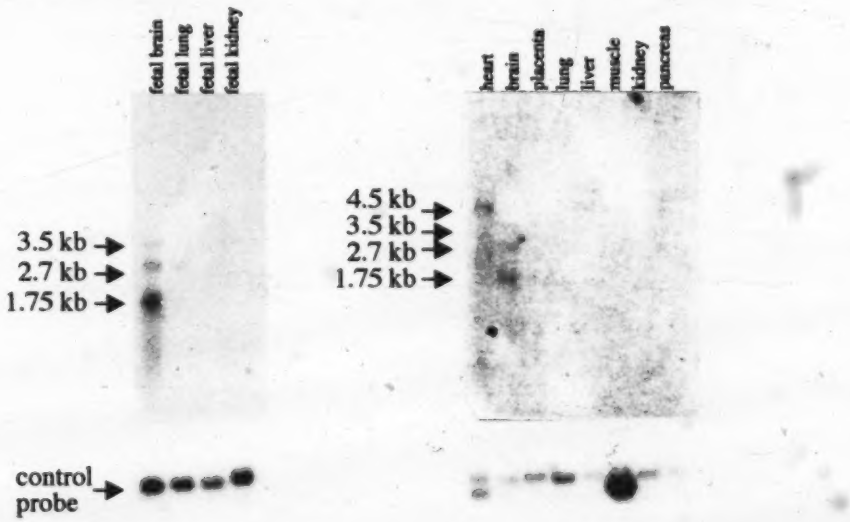
unfavourable for *SOX1*, contributing to the paucity of results for RT analysis. Another conundrum is the discrepancy between the 3.9 kb embryonic brain transcript described in the report by Malas and coworkers (1997) and the 4.6 kb transcript detected in fetal brain and other tissues by my Northern blot experiments. Further experiments were performed both with Northern blots and cDNA sources to elucidate the *SOX1* transcript map, and will be presented in the section "Transcript Mapping for *SOX1*" below.

The findings of high levels of *SOX1* expression in human fetal kidney as well as mouse kidney were intriguing. Sequence analysis revealed three binding sites early in the *SOX1* open reading frame for the transcription factor product of the Wilms tumour gene, *WT1*. The WT1 target sequence, GCGGGGGCG, is found beginning at *SOX1* positions 92, 104, and 122. *WT1* is a tumour suppressor gene mutated in patients with Wilms tumour, a childhood embryonal kidney tumour. These sequence observations along with the kidney expression findings fostered a sideline investigation of *SOX1* expression in relation to *WT1* with the hypothesis that *WT1* may be a repressor of *SOX1* transcription. In order to look for elevated *SOX1* expression in *WT1* negative tumours, a technician from the Wilms tumour research group (K. Dietrich) used the 450 bp *Pst* I – *Sst* I fragment of *SOX1* as a probe on a Northern panel derived from 16 Wilms tumours. A band at approximately 4 kb band was detected in all of the tumours, confirming *SOX1* expression in these tissues. However, a parallel hybridization with *WT1* in order to establish a relationship between the expression level of the two genes was not completed.

2. *SOX2*

Expression analysis for *SOX2* was limited to multiple tissue Northern (Clontech) hybridizations. A previous report of *SOX2* expression had detected a 3.5 kb transcript in NTera2 cells (Stevanovic et al. 1994). I used the entire clone B from the fetal brain library screening as a Northern probe. Transcripts were detected primarily in fetal brain and adult brain. Three bands were evident in both fetal and adult brain, sized at approximately 3.5, 2.7, and 1.75 kb, with intensity increasing from largest to smallest (**Figure 20**). Longer exposure revealed additional bands at 4.5 kb in adult heart and 2.7 kb in fetal lung. The 3.5 kb bands match the transcript size reported in NTera2 cells, and brain-specific expression fits the proposed role for *SOX2* in neural development (Stevanovic et al. 1994).

Figure 20. *SOX2* expression profile on Northerns with clone B probe



Because clone B was subsequently found to be chimeric, there is a possibility that some or all of the bands detected with the clone B probe represent transcripts

from the *NSP* gene. The *NSP* gene, on chromosome 14, is transcribed from alternate promoters to produce transcripts of 3.4, 2.3, and 1.8 kb. The 3.4 kb transcript is present in brain, the 2.3 kb transcript only in one small-cell lung cancer cell line (NCI-H82), and the 1.8 kb transcript in a variety of small-cell lung cancer cell lines (van de Velde et al. 1994). The *NSP* gene matches only 200 bp of end sequence on clone B, at which point the end sequence abruptly diverges from the *NSP* sequence, suggesting that the remaining 1.5 kb of clone B are not from chromosome 14. Since the majority of clone B likely represents *SOX2*-derived sequence, the Northern results may be plausibly interpreted as indicative of the *SOX2* expression profile. Unequivocal results would require probing with a fragment known to contain only *SOX2*.

E. Transcript Mapping for *SOX1*

The absence of transcription initiation and termination signals in the 2.2 kb of *SOX1* DNA sequence determined by subcloning and sequencing the genomic P1 DNA implied that the untranslated regions of *SOX1* extend further in both 5' and 3' directions. Since Northern blot analysis identified a 4.6 kb band for the *SOX1* transcript (**Figure 19**), there is likely approximately 2 kb of additional untranslated sequence in the transcript. I used a number of strategies to attempt to elucidate the structure of the complete *SOX1* transcript, including primer extension, RACE, RT-PCR, Northern blot differential analysis, additional library screening for cDNA clones, and *in silico* gene feature modelling. Delineation of the transcript was intended in part as a prelude to promoter analysis for *SOX1*; however, because of the difficulties of transcript mapping, promoter analysis was not carried out.

1. Primer extension

Primer extension reactions are designed to determine the transcription start site (TSS) in the 5' untranslated region (UTR) of a gene. A reverse primer presumed to be close to the TSS is hybridized to an RNA population which expresses the gene of interest. Reverse transcriptase (RT) added to the reaction then extends a DNA copy of the RNA until it reaches the 3' end of the RNA, which should correspond to the TSS. The extension product is then run on PAGE alongside of a sequencing reaction carried out using the same primer to determine the precise location of transcription initiation.

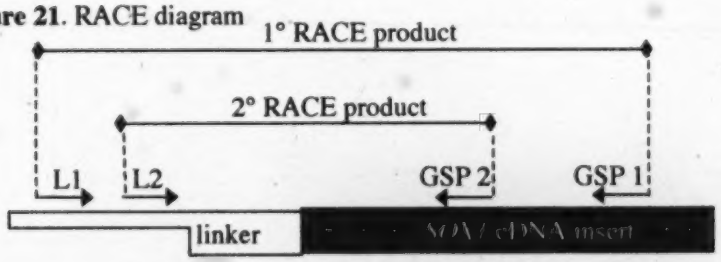
Primer extension reactions for *SOX1* were performed using the following combinations of primers and RNA sources: primers 221 and 279 on lymphoblast RNA, and primers 64 and 100 on NTera2 RNA. A substantial product band was detected from the reactions with the lymphoblast RNA; however, the band migrated at a constant distance of about 20 bp from the primer for each of the primers. Since the two primers used are separated on the *SOX1* sequence by a distance of about 60 bp, the products should have also differed in size by about 60 bp (see **Appendix C** for primer sequences and locations). The primer extension reactions carried out for the NTera2 RNA failed to produce a notable product band. Unlike primer extension, the other methods used for transcript mapping were designed to obtain both 3' and 5' UTR sequences.

2. RACE and RT-PCR

Numerous RACE PCR reactions were carried out on various cDNA sources in an attempt to amplify both 5' and 3' UTR sequences for *SOX1*. All of the cDNA sources were Marathon-ready cDNAs from Clontech. These are basically

unamplified libraries of linker-ligated cDNA designed to provide excellent templates for RACE reactions. A simple diagram of the RACE method for 5' amplification is depicted in **Figure 21**. Linker-specific primers (L1 and L2) are paired with gene-specific primers (GSP1 and GSP2) to selectively amplify the end of a cDNA clone. The linkers are designed with single stranded ends so that in the first round of amplification, only the gene specific primers will proceed. The primary RACE product is derived from the outermost primer pair, L1 and GSP1. A secondary RACE product is then amplified from the primary product by a nested PCR reaction using the inner primer pair, L2 and GSP2. This secondary product is then cloned using a TA cloning kit and analyzed by digest, PCR, or sequencing to reveal its identity.

Figure 21. RACE diagram



Legend: L1, L2 linker-specific primers
GSP 1, GSP 2 gene-specific primers

Each cDNA source was tested by PCR with *SOX1* primers as a positive control before proceeding with RACE reactions. The linker-specific primers used were AP1 (outer) and AP2 (nested). A variety of gene-specific primers were used. Unfortunately, no *SOX1* cDNA ends were successfully amplified using this method, although many products were cloned and sequenced. The reason for this failure may again lie in the GC-rich nature of the *SOX1* gene, which presents an

obstacle to PCR amplification. Alternatively, *SOX1* representation among the cDNA sources tested may have been below the necessary threshold level for detection. A third possibility is that the gene-specific primers selected for the RACE reactions were too far distant from the cDNA ends, making PCR amplification unfeasible. **Table 8** summarizes the primers and cDNA sources used for 5' and 3' RACE for *SOX1*.

Table 8. RACE reactions for *SOX1*

cDNA Source	Primary GSP	Secondary GSP
5' RACE		
Heart RACE-ready cDNA	279	221
Fetal brain RACE-ready cDNA	221	900B
	279	
	857	G, 221, 473
	1012	
	1259	1012
	64	
	100	100, 64
	364, 377	179
3' RACE		
Fetal brain RACE-ready cDNA	2138	506F
	1611	1809, 1889
	1889	1889, 1972
	367F	336F
	506F	367F

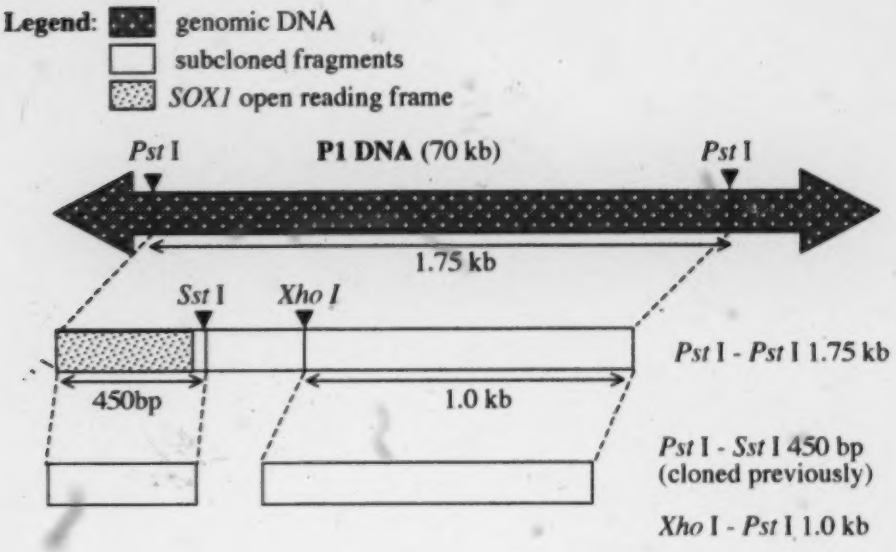
Mapping the 3' end of the *SOX1* transcript was augmented by the use of RT-PCR on RNA isolated from the NTera2 cell line. Three primer sets were used to test for detectable transcript (see **Appendix C** for primer sequences and locations). Primer pair 367F/R yielded a positive result, whereas primer pairs 506F/R and 359F/R were negative. Since the 367F/R pair is more 3' than the 506F/R or 359F/R pairs, these results may indicate the presence of an intron between 367F/R

and the main coding region of the *SOX1* transcript. Alternatively, the 367F/R primers may be amplifying a non-*SOX1* transcript.

3. Genomic cloning and differential Northern blot analysis

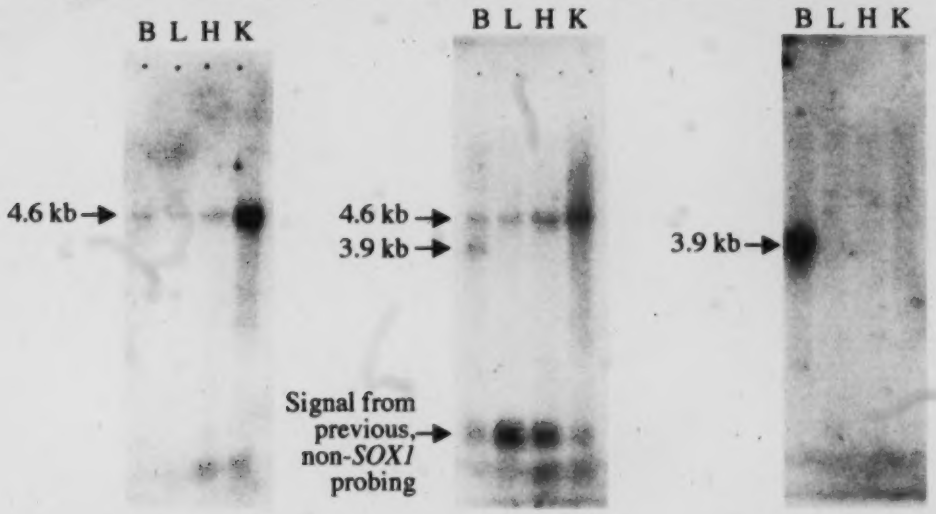
Additional subcloning of the *SOX1* genomic region was undertaken in order to generate suitable probes for differential Northern blot analysis and to create a contig that could be used as a basis for promoter analysis. From the 3' region of the *SOX1* sequence, a 1.75 kb *Pst* I fragment was cloned from the P1 DNA. This 1.75 kb subclone was identical to the probe used on Northern blots by Malas and coworkers, who reported a 3.9 kb transcript in fetal brain (1997). The 5' end of the 1.75 kb subclone was comprised of the 450 bp *Pst* I – *Sst* I fragment that I had used as a probe earlier to detect a 4.6 kb *SOX1* transcript in multiple tissues. The 3' end of the 1.75 kb *Pst* I subclone was in turn further subcloned to produce a 1.0 kb *Xho* I – *Pst* I subclone, which did not overlap with the 450 bp *Pst* I – *Sst* I fragment. The subcloning of the 3' region of *SOX1* is depicted in **Figure 22**.

Figure 22. Subcloning of *SOX1* 3' region



The 1.75 kb *Pst* I subclone and the1.0 kb *Xho* I – *Pst* I subclone were used in turn to probe a set of fetal and adult Northern blots. The results were unexpected and quite intriguing, especially in combination with the earlier result of detecting a 4.6 kb *SOX1* transcript in multiple tissues by using the 450 bp *Pst* I – *Sst* I fragment as a probe. Probing with the full length 1.75 kb *Pst* I subclone again detected a widespread 4.6 kb transcript, but also detected a 3.9 kb transcript specific to fetal brain. Probing with the1.0 kb *Xho* I – *Pst* I subclone detected only the 3.9 kb transcript in fetal brain, and nothing in other tissues. The autoradiographs of the human fetal Northerns hybridized with the three probes are shown in **Figure 23.** These results are also indicated on **Figure 25**, *SOX1* genomic map, where the three probes are labelled **A** (450 bp *Pst* I – *Sst* I), **B** (the1.0 kb *Xho* I – *Pst* I), and **C** (1.75 kb *Pst* I).

Figure 23. Differential results with *SOX1* probes on human fetal Northerns



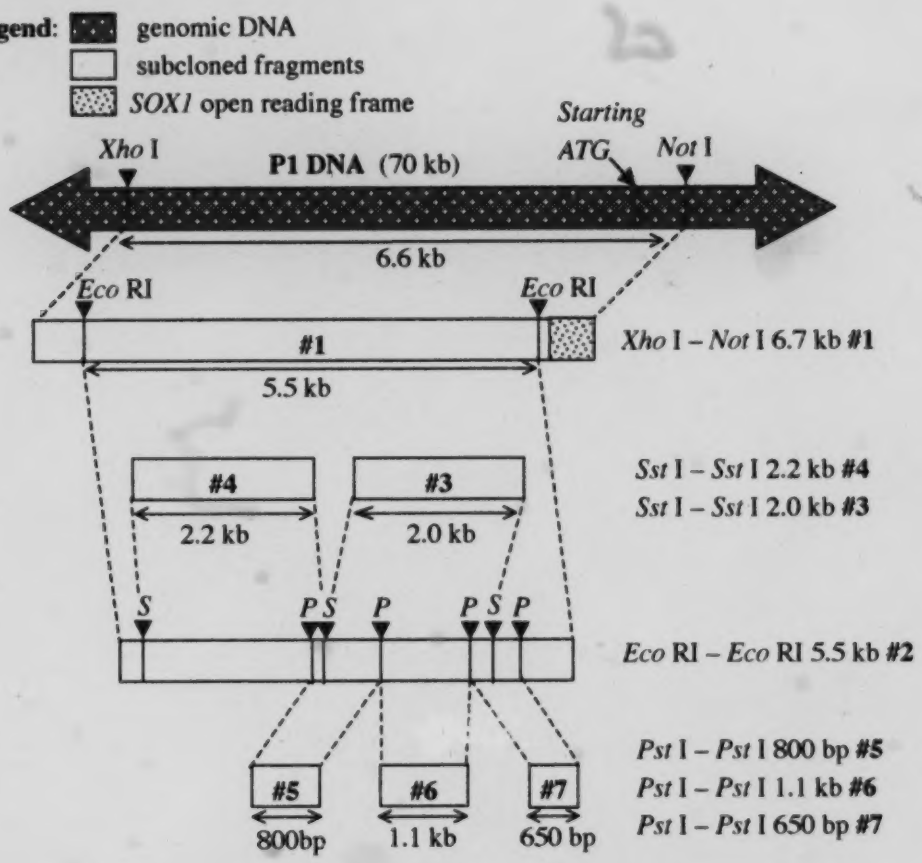
Probe: 450 bp *Pst* I – *Sst* I 1.75 kb *Pst* I 1.0 kb *Xho* I – *Pst* I

Label: **A** **C** **B**

Legend: B, fetal brain L, fetal lung H, fetal liver K, fetal kidney

The 5' UTR of *SOX1* was also subcloned from P1 genomic DNA. A series of digestions and clonings with the enzymes *Not* I, *Xho* I, *Eco* RI, *Pst* I, and *Sst* I produced subclones numbers 1 through 7, indicated diagramatically in **Figure 24**. These subclones were partially sequenced at the sequencing facility at the University of Victoria in order to analyze the region for gene features such as the TSS and splice sites.

Figure 24. Subcloning of *SOX1* 5' UTR



Four of the seven subclones from the 5' region were used as probes to test multiple tissue Northern blots (**Figure 25**). These hybridizations were intended to identify the 5' boundary of the *SOX1* transcript because probes which lay upstream of the transcript would not be expected to detect any bands. No hybridization was found for probes #3, 5, and 6. Probing with #7 produced the widespread signal at 4.6 kb, indicating that the 5' end of the transcript lay within the 650 bp of #7. To further refine this localization, the *Pst* I – *Pst* I 650 bp #7 clone was subdivided with the enzyme *Sst* II into a 5' 400 bp fragment and a 3' 250 bp fragment. These fragments were cloned to also use as Northern probes.

The 400 bp 5' *Pst* I – *Sst* II fragment (#7A) also detected the 4.6 kb band on Northern blots. These differential Northern hybridizations suggested that the 5' end of the *SOX1* transcript lies in a small area of about 200 bp contained within the *Pst* I – *Pst* I 650 bp subclone #7 (between the 3' end of the *Sst* I – *Sst* I 2.0 kb subclone #3 and the *Sst* II site). There is a possibility, however, of another 5' exon lying further upstream of the transcript end detected by the Northern probings.

An additional probing of the Northern blots was performed to confirm the presence of the *SOX1* HMG box within the transcript. A 487 bp *Eco* RI – *Sca* I fragment which contained the entire HMG box was isolated from the *Eco* RI – *Pst* I 850 bp subclone encompassing the 5' half of *SOX1* (see **Figure 14**). The 4.6 kb transcript was detected with this probe as well (**Figure 25**).

A scale diagram of the *SOX1* genomic region is presented in **Figure 25**. The various subclones from the 5' and 3' regions are indicated, along with the results obtained by using them as probes on multiple tissue Northern blots.

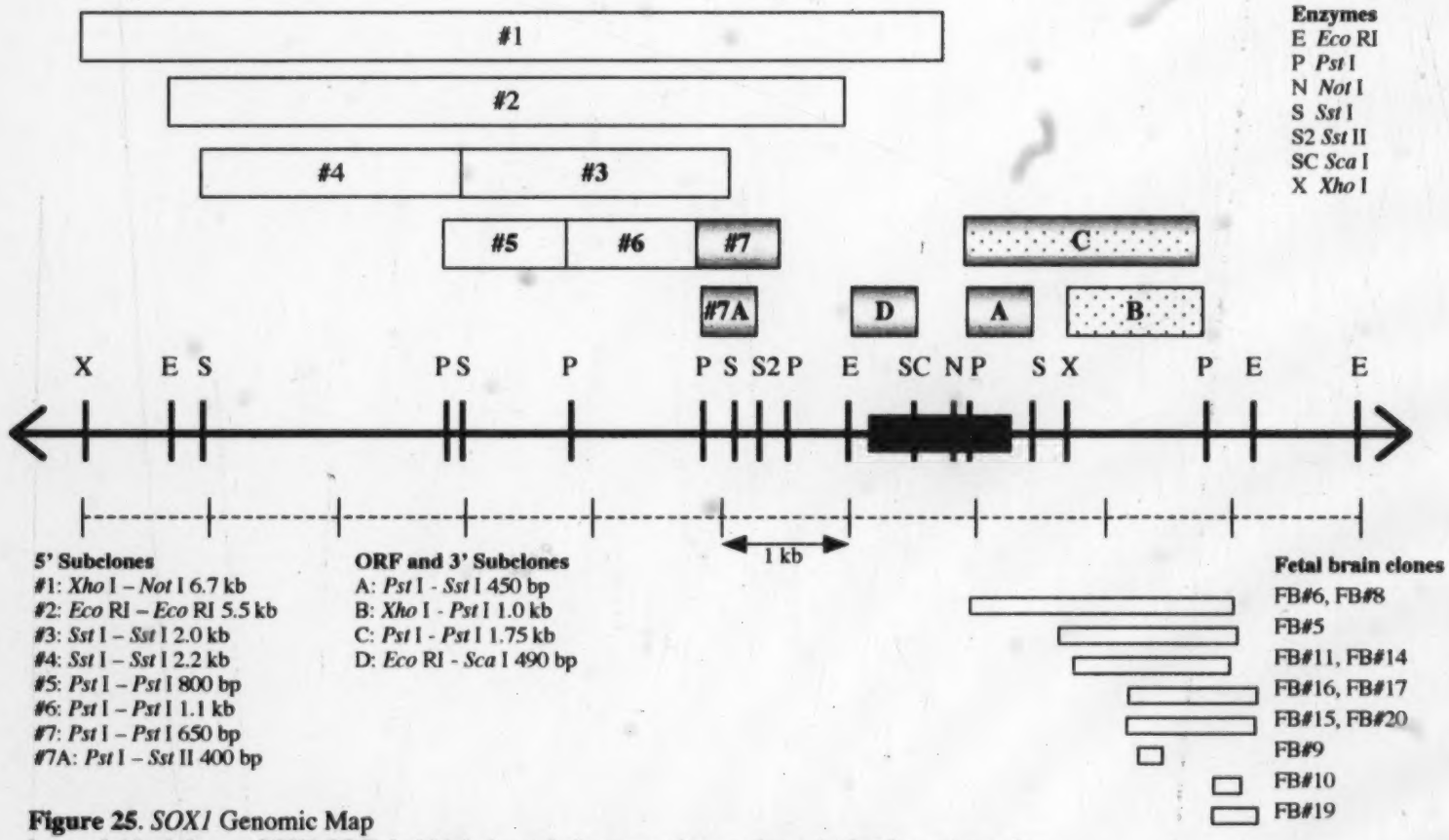


Figure 25. *SOX1* Genomic Map
Legend: black box = *SOX1* ORF; bold label = subclones used as probes; shaded boxes = probes detecting 4.6 kb transcript; speckled boxes = probes detecting 3.9 kb transcript

4. Fetal brain library screening

A second fetal brain cDNA library screening was undertaken in order to obtain cDNA clones representing the 3' and perhaps 5' ends of the *SOX1* transcript. The 1.75 kb *Pst I* clone of the 3' region of *SOX1* was used as a probe to screen approximately 600 000 plaques. This probe was chosen because it would theoretically detect transcripts of both the 3.9 kb and 4.6 kb species. Out of roughly 100 positive plaques, 20 were picked for further analysis. Secondary screening was conducted with two probes, designed to detect the 3.9 kb and 4.6 kb transcripts separately, hybridized to duplicate filter lifts of the secondary platings. Excessive background made it impossible to detect genuine positives from the probing with the 450 bp *Pst* I – *Sst* I fragment of the *SOX1* gene, designed to detect the 4.6 kb transcript. Fourteen positives were selected to carry forward to tertiary screening from probing with the1.0 kb *Xho* I – *Pst* I fragment, intended to detect the 3.9 kb transcript. These fourteen positives were subsequently converted to plasmid form using the Stratagene excision protocol for the λ-ZAP phagemid.

Results from the restriction mapping and sequencing of these fourteen clones are presented in **Table 9**. A scale diagram of the *SOX1* genomic region is presented in **Figure 25**. In addition to the various 5' and 3' subclones, the *SOX1* clones recovered by screening the fetal brain library are depicted to indicate approximate size and location in the genomic sequence.

Table 9. Clone analysis from fetal brain cDNA library

Clone Number	Insert size (kb)	Composition	Region of *SOX1* represented by clone	Comment
FB#3	3.4	chimeric	none	
FB#5	1.6	chimeric	1373-2763	
FB#6	3.4	chimeric	682-2736	Identical to FB#8; 94bp gap in *SOX1*
FB#7	2.85	chimeric	none	
FB#8	3.4	chimeric	682-2736	Identical to FB#6; 94bp gap in *SOX1*
FB#9	2.85	chimeric	1816-2002	
FB#10	3.1	chimeric	2512-2728	
FB#11	1.2	non-chimeric	1389-2603	Identical to FB#14
FB#14	1.2	non-chimeric	1389-2603	Identical to FB#11
FB#15	1.0	non-chimeric	1810-2814	Identical to FB#20
FB#16	1.25	chimeric	1804-2818	Identical to FB#17
FB#17	1.25	chimeric	1804-2818	Identical to FB#16
FB#19	2.3	chimeric	2823-2475	
FB#20	1.0	non-chimeric	1810-2814	Identical to FB#15

A total of 10 unique clones was isolated from this second fetal brain library screening and four clones were duplicates. Eight of the 10 unique clones represented *SOX1*, of which 6 were chimeric, containing sequence from other cDNAs, and two contained exclusively *SOX1* sequence. Most of these *SOX1*-containing clones mapped to the region covered by the 1.0 kb *Xho* I – *Pst* I probe, which is not surprising since they were identified using this probe. Only one of the eight new *SOX1* cDNAs (FB#6, 8) overlapped the *SOX1* open reading frame (positions 1 – 1173), and it contained a 94 bp deletion. This deletion did not appear to be a good candidate for an intron because the genomic sequence lacks consensus splice sites in this region.

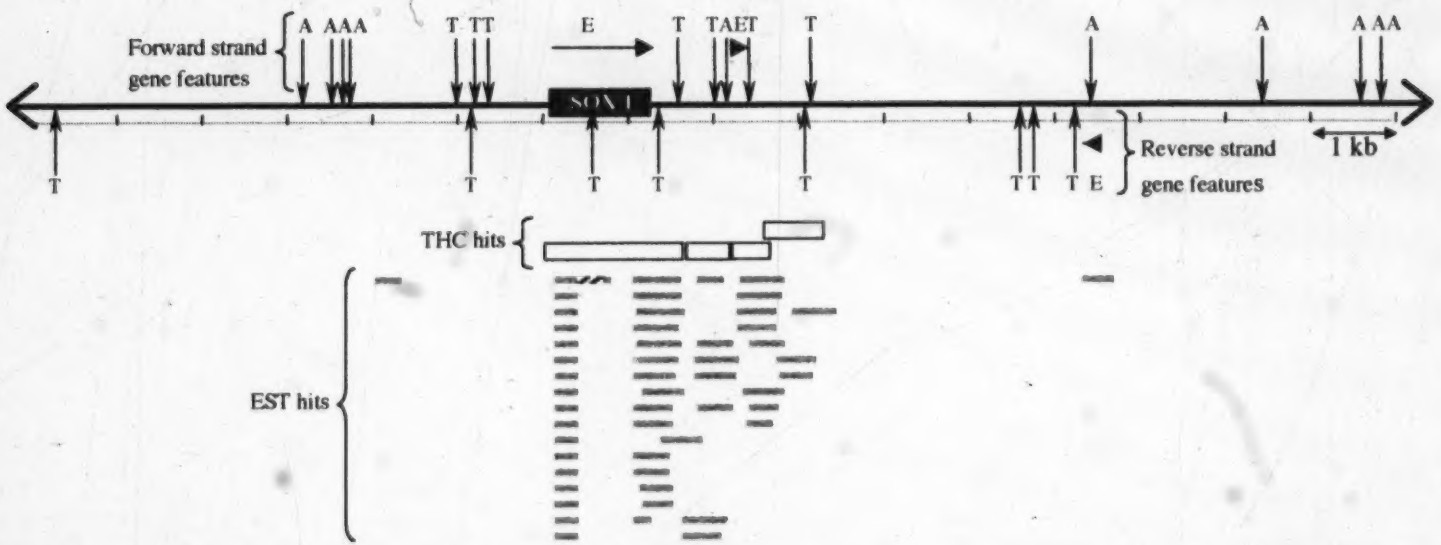
5. *In silico* gene feature modelling

In order to help interpret the experimental data, and to make testable predictions, regular analysis of the *SOX1* genomic sequence was undertaken using various online programs. Only the most recent analysis will be considered for the sake of brevity and relevance. For the purpose of pinpointing the ends of the transcript and identifying any alternate transcript forms, searches were performed using genomic sequence to predict the transcription start site (TSS), intron/exon structure, and polyadenylation signal for *SOX1*. Regular searches were also performed for expressed sequence tags (ESTs) in the *SOX1* genomic region. ESTs represent sequence which has been found to be expressed as mRNA, in other words, sequence derived from a gene. The detection of ESTs can thus indicate which other genes or exons lie within the vicinity of *SOX1*, or even help identify upstream or downstream exons belonging to *SOX1*. The most recent searches were carried out using approximately 16 kb of genomic sequence, 6 kb upstream and 9 kb downstream of the *SOX1* ORF. This genomic sequence was generated

by combining two pieces of the genomic sequencing contig RP11-310D8 based on their matches to either end of the *SOX1* gene.

I used a number of different gene feature prediction programs to analyze approximately 16 kb of the *SOX1* genomic sequence, and looked for points of agreement among the predictions as indications of more probable structures (URL addresses are given in Materials and Methods). For both directions of read, the first A of the initiation codon is taken as position number one and the last C before the termination codon is position number 1173. Using this numbering system, the genomic region analyzed extends from –6218 to +10493 bp on the forward strand (which contains *SOX1*) and from +10493 to –6218 on the reverse strand. **Figure 26** summarizes predictions for the transcription start site (TSS), intron/exon structure, and polyadenylation signal for *SOX1* that are supported by two or more prediction programs within 10 – 100 bp. The results of BLAST searches with this 16 kb genomic section for human ESTs and THCs are also indicated in **Figure 26**.

The results of the web-based prediction analysis for the *SOX1* genomic region were not highly conclusive. There was not a high degree of concordance between the various different programs, suggesting that the region does not have a simple gene structure. Several proposals will be put forward in the discussion to try and fit the transcript mapping predictions of the *in silico* analysis with the experimental data.



EST hits (consolidated)	THC hits	Transcription Start Sites (T)	Predicted Exons (E)	Polyadenylation Sites (A)
-1966 – -1665, 144 – 799, 1081 – 2289, 2306 – 3458, 6364 – 6715	1 – 1670, 1672 – 2215, 2388 – 2830, 2759 – 3458	Forward: -988, -758, -687, 1649, 2172, 2466-2471, 3182 Reverse: -5807, -902, 604, 1191, 3040, 5635, 5786, 6138	Forward: -40 – 1176, 2344 – 2392 Reverse: 6693 – 6725	Forward: -2895, -2476, -2326, -2283, 2196, 6657, 8423, 9682, 9937, 9995 Reverse: none

Figure 26. *In silico* predictions of gene features, ESTs, and THCs for the *SOX1* genomic region
Legend: T, transcription start site; **E,** predicted exon; **A,** polyadenylation site; black box *SOX1* ORF

F. Mutation Studies

1. *SOX1*

Several factors suggested that a cataract patient screen for mutations in *SOX1* would be appropriate. First, mutations in other *SOX* genes have been found in association with human disorders. Second, its role in crystallin gene expression in animal models suggested that *SOX1* mutations could disrupt crystallin production in human eyes and potentially cause cataracts. Finally, a mouse deleted for *Sox1* had recently been produced which exhibited cataracts and microphthalmia (Nishiguchi et al. 1998). With the help of Dr. E. Héon (Eye Research Institute of Canada, University of Toronto) and Dr. I.M. MacDonald (Department of Ophthalmology, University of Alberta), we collected DNA from three families and 11 individual patients with cataracts and microphthalmia and screened this panel of 18 individuals for *SOX1* mutations by single stranded conformation polymorphism (SSCP) analysis and DNA sequencing.

Developing primers to effectively analyze *SOX1* by SSCP and sequencing was a challenge because of the very high GC content of the gene. A total of six primer pairs were devised that spanned the coding region of *SOX1* (**Table 4** in M & M, **Figure 27**). These primers were used to amplify DNA segments from 18 individuals belonging to families with microphthalmia and cataracts. Twenty-one individuals with other anterior segment anomalies, including aniridia and Peter anomaly, were also tested with two of the primer sets, 582 + 857 and 1043 + 1241, which spanned the HMG box and the PRD-type repeat respectively. Patient phenotypes are described in **Table 6** in Materials and Methods.

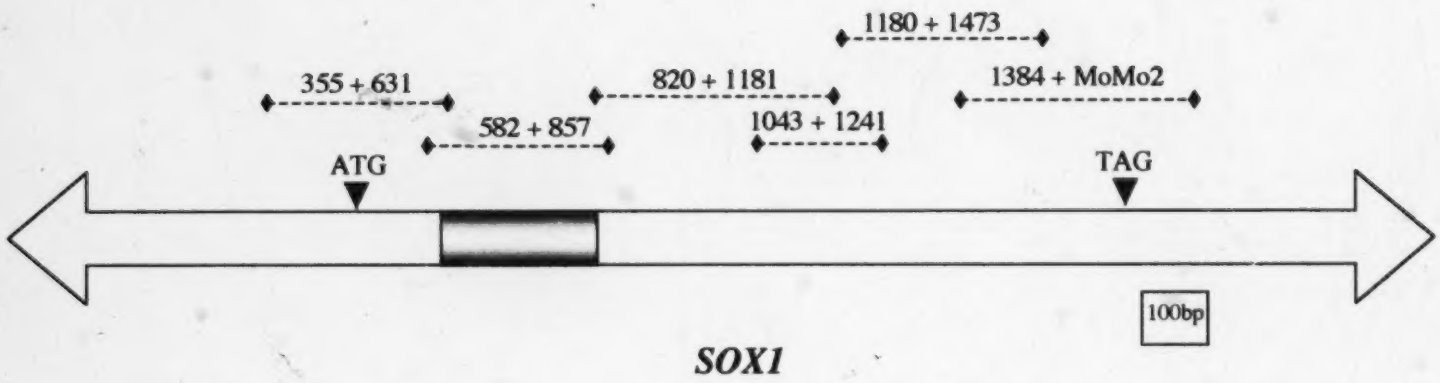
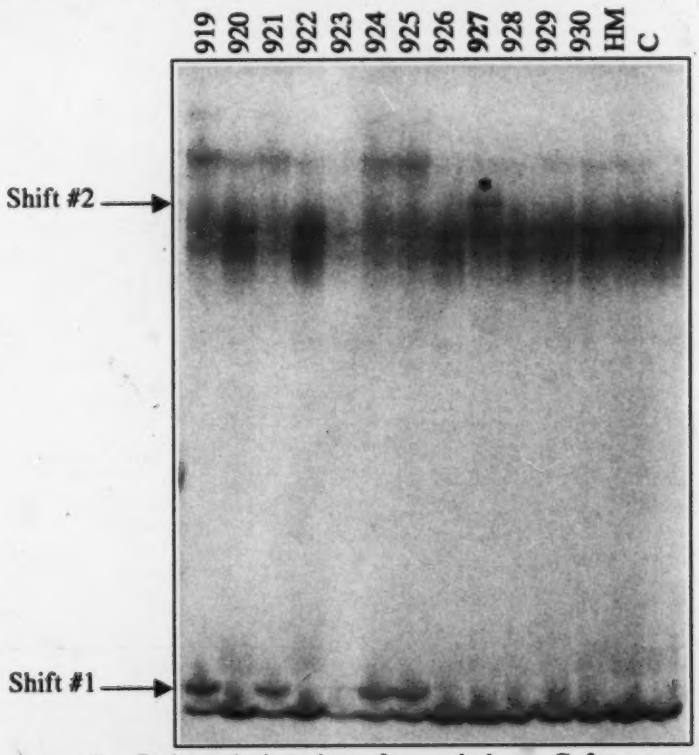


Figure 27. *SOX1* SSCP PCR contig. HMG box domain is shaded. Start and stop codons are indicated with arrows.

No alterations were detected by SSCP analysis for five of the primer sets in the individuals tested. Two shifted-band patterns were detected for the primers 1384 – MoMo2 among the 18 individuals in families with microphthalmia and cataracts (**Figure 28**). The first shifted-band pattern (#1) was found in four of 18 individuals, and the second shifted pattern (#2A) was found in only one person, patient #927. Subsequently, DNA from 63 unaffected individuals of various ethnicity (126 chromosomes) was tested by SSCP analysis with primers 1384 – MoMo2. This control screening revealed 13 shifts matching the shift #1, two shifts resembling shift #2A, and one new shift (#3).

Figure 28: SSCP analysis for 1384 – MoMo2 fragment



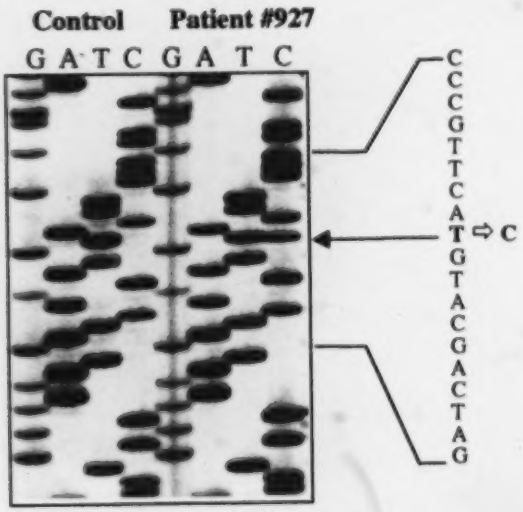
Legend: Patient designations for each lane, C for control individual. Patient #927 is boldfaced, and his shift (#2) marked by asterisk.

DNA sequencing was carried out on PCR fragments amplified from appropriate individuals to ascertain the nature of the three alterations detected by SSCP analysis in the 1384 – MoMo2 segment. The first shift (#1) was found to correspond to a 6 nucleotide duplication of the sequence GGCGTG at position 1146 of the *SOX1* coding region (autoradiograph not shown). This change is predicted to duplicate a glycine and a valine residue in the SOX1 protein, but will leave the reading frame unaffected. This variation occurs near the C-terminus of the protein, as termination would occur just ten amino acids downstream of the duplication. This alteration was found to be present in the control individuals without cataracts at a frequency of 10.9% and therefore can be considered to be a benign polymorphism.

The second SSCP shift (#2) detected for fragment 1384 – MoMo2 turned out to be rather complex. After sequencing, a heterozygous transition of T to C was found at position 1039 of the *SOX1* coding region in patient #927 (**Figure 29**). This substitution would be predicted to replace the normal tyrosine (TAC) at amino acid position 347 with a histidine (CAC). The affected residue would lie between the last two polyalanine stretches of SOX1. When DNA products were sequenced from the two unaffected individuals who apparently had a similar SSCP shift to patient #927, no sequence changes were found. To distinguish the two alleles which have the same apparent SSCP shift, the T1039C shift was designated #2A, and the control individuals' shift called #2B. Although there is presumably a change in sequence which has escaped detection in these two control individuals with shift #2B, the T at position 1039 was clearly not changed, suggesting that the second shift detected in the control individuals represents

another sequence polymorphism, different from the T1039C transition. The lack of alteration at position 1039 in the control chromosomes was subsequently confirmed by *Rsa* I digestion (see below). The frequency of this allele for shift #2B was calculated to be 1.6% among the control chromosomes tested (i.e. 2 out of 126 chromosomes). Alternatively, the second shift in the control individuals may be a PCR artifact which does not represent a sequence change.

Figure 29. DNA sequence showing T to C transition for patient #927



DNA from the control individual with the third shift (#3, not shown) was also sequenced for the 1384 – MoMo2 region; however, once again no change in sequence was detected. Like the second shift in the control individuals, there is likely a change in sequence which has escaped detection and is assumed to be benign, or the shift may represent a PCR artifact. The frequency of this third shifted allele was calculated at 0.8% of control chromosomes, but it was only detected in a single individual, reducing the level of confidence in the frequency

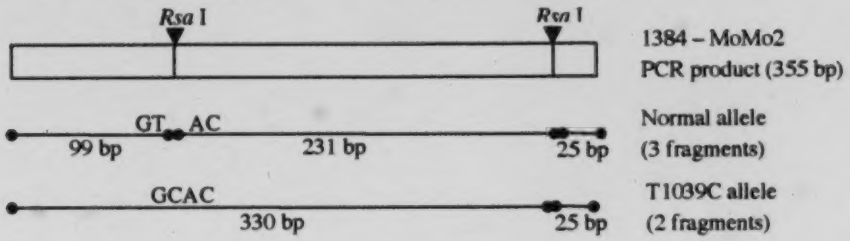
calculation. **Table 10** summarizes the alterations found in the *SOX1* gene by SSCP, with sequencing results when determined.

Table 10. *SOX1* SSCP alterations detected within 1384 – MoMo2 fragment

SSCP Alteration	DNA sequence change	Amino acid change	Frequency
1	GGGCGT duplication at position 1146	Extra Gly, Val at position 383	13/126, or 10.3% of control chromosomes
2A	T ⇨ C at position 1039	Tyr ⇨ His at position 347	1/14, or 7% of patient families 0/126 of control chromosomes
2B	Unknown	None inferred	2/126, or 1.6% of control chromosomes
3	Unknown	None inferred	1/126, or 0.8% of control chromosomes

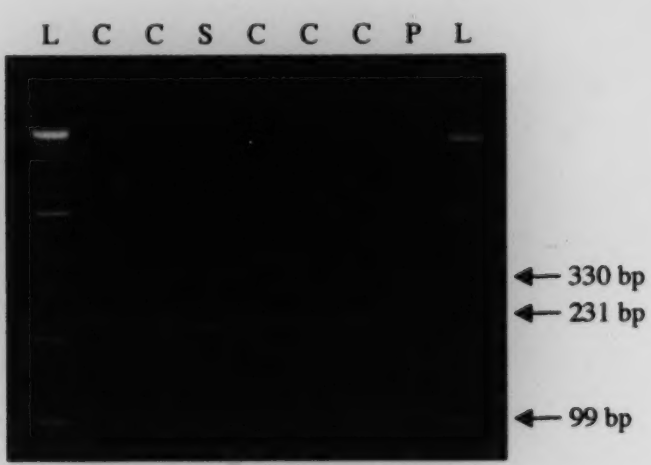
The T to C transition found for patient #927 was predicted to abolish a GTAC restriction site for *Rsa* I in the *SOX1* sequence. This change was used to confirm the sequence change with a simple restriction digestion of the 1384 – MoMo2 PCR product. Normal individuals are expected to have three fragments, sized 99, 231, and 25 bp, after *Rsa I* digestion of the 355 bp PCR product. Individuals carrying the T1039C transition should show a larger, uncut fragment at 330 bp. **Figure 30** diagrams the *Rsa* I sites expected in the 1384 – MoMo2 PCR products.

Figure 30. *Rsa* I map of the 1384 – MoMo2 PCR product



Rsa I digestion of the 1384 – MoMo2 PCR product was carried out for patient #927 (shift #2A), a control individual with the upper shift on SSCP (shift #2B), and several unshifted individuals. The band pattern observed by electrophoresis matched the predicted pattern and confirmed that patient #927 had a unique alteration. A gel photo is depicted in **Figure 31**. The 25 bp band cannot be seen.

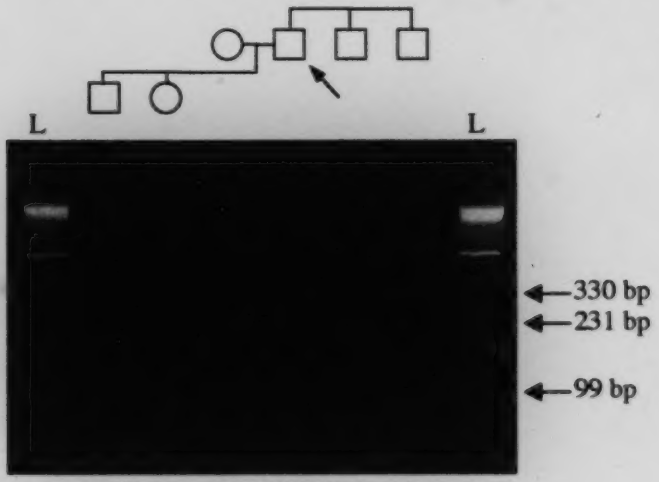
Figure 31. *Rsa I* digest of 1384 + MoMo2 PCR products for patient panel



Legend: C, control individual S, upper shift control individual
P, patient #927 L, 100 bp ladder

The diagnosis for patient #927 was zonular cataracts with no microphthalmia detected. Interestingly, this patient's spouse had microphthalmia and cataracts, but was not part of the initial screening. After finding an alteration in the *SOX1* sequence of patient #927, DNA was obtained via Dr. E. Héon in Toronto for additional family members, including his wife, two children, and two brothers. PCR products for 1384 – MoMo2 from the new samples were sequenced, and the son and one brother were found to have the T to C transition detected in patient #927. These alterations were confirmed by *Rsa* I digestion. The results for the digest assessment of the family is shown in **Figure 32**.

Figure 32. *Rsa I* digest of 1384 + MoMo2 PCR products for patient #927 family



Legend: Arrow indicates patient #927. L, 100 bp ladder

Clinical information for the family is presently very sketchy. Patient #927, his wife, and both of their children have cataracts. The mother and children also have microphthalmia, and the children have secondary glaucoma. No information about the clinical status of the two brothers of patient #927 is available. Final

interpretation of the T1039C *SOX1* alteration must therefore await the arrival of further clinical reports.

2. *SOX9*

A child with campomelic dysplasia (CD) from a family in northern Alberta was seen by Dr. J.S. Bamforth (Department of Medical Genetics) at the genetics clinic of the University of Alberta. Since this patient had survived longer than average for CD, the clinic was interested in confirming the diagnosis of CD by finding a causative mutation in the patient's *SOX9* gene. A project student (J. Giordano) carried out the task of screening the patient's *SOX9* gene by SSCP and DNA sequencing.

Primers for *SOX9* SSCP analysis were chosen from those published by Kwok et al. (1995) (**Figure 33**). An SSCP shift was observed for the patient in the 3' end of the *SOX 9* ORF amplified by primer pair Y and W (**Figure 34A**). DNA sequencing demonstrated a +1 frameshift mutation in the patient, resulting from the insertion of a single guanine nucleotide in a stretch of four guanines. The four guanines are nucleotides 1453-1456, and the mutation is designated ins1456G, with numbering starting from the initiating ATG (**Figure 34B**). The patient is heterozygous for the ins1456G mutation and carries a normal allele of the *SOX9* gene as well (**Figure 34B**).

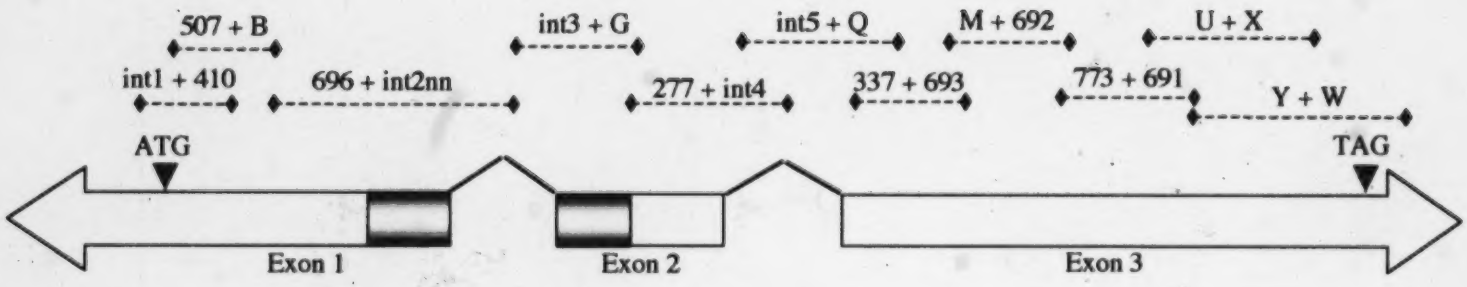


Figure 33. *SOX9* SSCP PCR contig. HMG box domain is shaded. Start and stop codons are indicated with arrows (based on Kwok et al. 1995).

The mutation creates a mutant *SOX 9* open reading frame that is 201 nucleotides longer than normal. The amino acid sequence is altered by the frameshift beginning at position 485, and the putative mutant SOX9 protein is 576 amino acids in length, 67 amino acids longer than the normal protein (depicted in **Figure 39** in the Discussion). The last 24 C-terminal amino acids of the normal SOX9 protein have been replaced by 91 mutant amino acids in the mutant SOX9 protein.

Only one previously described *SOX9* mutation, an insertion of 4 base pairs at nucleotide position 1519 described by Kwok et al. (1995), is further 3' than the mutation detected in our CD patient (see **Figure 39** in the Discussion).

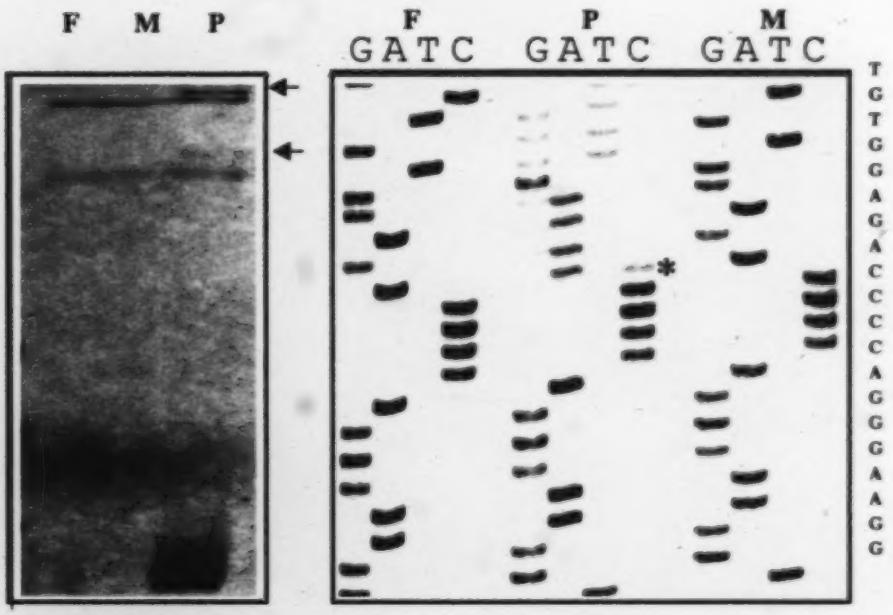


Figure 34 A. *SOX9* SSCP shifts for primers Y+W indicated by arrows for patient. **B.** *SOX9* sequence of patient and parents showing reverse strand sequence at right with position of insertion indicated by asterisk. (F, Father; M, Mother; P, Patient)

DISCUSSION

A. The *SOX1* gene

The major focus of my work has been the cloning and characterization of the human *SOX1* gene. After unsuccessfully trying to clone *SOX1* from a human fetal brain cDNA library, I selectively amplified it the way many members of the *SOX* family were initially identified, by PCR of the HMG box region (Gubbay et al. 1990). Using PCR to obtain a genomic clone of the *SOX1* region set the stage for further sequencing, localization, and expression studies. Because the *SOX1* ORF is intronless, a cDNA was not required to determine sequence of the coding region or size of the gene.

The very high GC content of the *SOX1* coding region (76%) and 1 kb of the upstream region (74%) made it very difficult to sequence and to design suitable PCR primers. Nonetheless, using stringent conditions I eventually assembled over 2 kb of sequence that I was confident represented the human *SOX1* gene. I concluded that the sequence which I established was the *SOX1* gene based on comparison with the known murine *Sox1* gene. The human and murine DNA sequences are 95% identical, and the protein sequences are 99% identical over the entire coding region, with a 100% amino acid match within the HMG box (**Figures 35, 38**). There are only four amino acid differences between human SOX1 and mouse Sox1 along the entire protein (indicated on **Figure 35**), and one of these substitutions is conservative. As well, I determined that human *SOX1* was located on chromosome 13q34, in a region that shares conserved synteny with the region of mouse chromosome 8 to which murine *Sox1* had been localized (Koizumi et al. 1995).

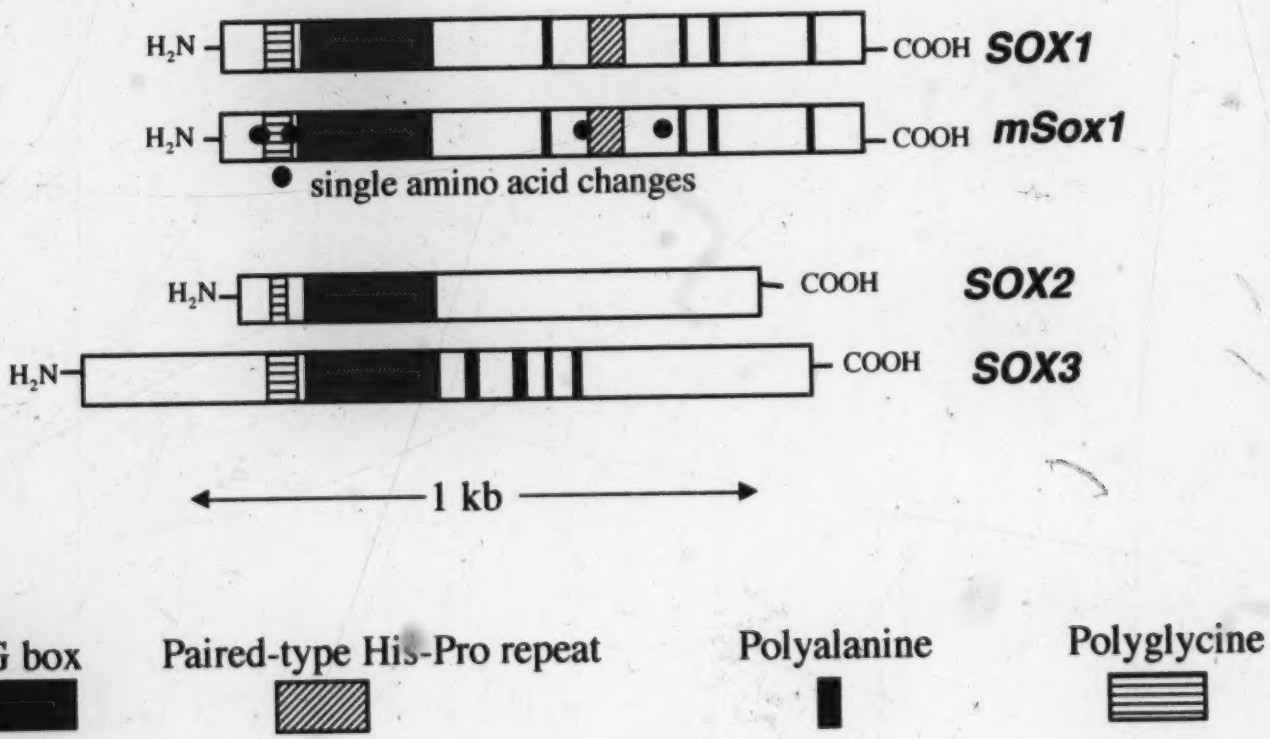


Figure 35. Comparative structural alignment for SOX1, Sox1, SOX2, and SOX3.

The absence of potential splice sites within the *SOX1* ORF suggests that it is intronless, like murine *Sox1*, human *SOX2*, *SOX3*, and other *SOX* group B genes. Within the HMG box, the SOX1 predicted protein is 96% identical to SOX2 and 91% identical to SOX3. Outside of the box, the identity drops to 40% for SOX2 and 46% for SOX3. However, SOX1 has some structural features in common with its closest B group relatives outside of the HMG box, including a polyglycine tract before the box, found in SOX1, 2, and 3, and several polyalanine stretches after the box, found in SOX1 and SOX3 (**Figure 35**, see also **Figure 38**). The high level of sequence conservation, both between human and mouse *SOX1* orthologues, as well as among the members of the human group B *SOX* genes, points to an important role for these genes in mammalian development.

The HMG domains of most *SOX* genes examined to date have been found to bind the consensus target sequence A/T A/T C A A A G (Harley et al. 1994; Mertin et al. 1999). Target binding sequences for Sox1 have been identified within crystallin gene enhancer regions by mutation and gel shift mobility assays in both chick and mouse systems. In the chick DC5 enhancer, Sox1 binds the target AACAATG, a close match to the consensus sequence (Kamachi et al. 1999). In mice, Sox1 binds a site present in the promoter of all γ-crystallin genes, namely ACAAA, which matches the core of the SOX consensus sequence (Nishiguchi et al. 1998). The crystallin enhancers in chick and mouse both carry the Sox1 target sequence on the opposite strand to the one that is transcribed.

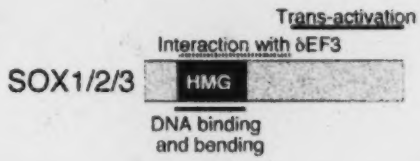
Sequence analysis suggests that several regions of *SOX1* are likely to be involved in the control of transcription. These transcriptional control regions include: a

polyglycine region upstream of the 79 amino acid HMG box, four polyalanine stretches downstream of the box, and a His-Pro PRD-type repeat region (**Figure** [redacted] Transcriptional repression has been associated with alanine-rich regions in other development control genes such as *engrailed* (Triezenberg 1995). As well, the *SOX1* gene is the only known *SOX* gene to contain a His-Pro rich PRD-type repeat, a 21 amino acid domain found in the Drosophila *paired* gene and related homeotic genes including *bicoid.* Although the function of this motif is not certain, chimeric proteins fusing the PRD repeat to the DNA-binding domain of Sp1 (a zinc finger protein) were able to activate a CAT reporter construct in *Drosophila* Schneider-3 cells (Cai et al. 1994). These gene fusion studies as well as other deletion experiments suggest that the PRD repeat is likely part of a transcriptional activation domain.

Sox1-specific transactivation studies in chick lens culture using sequential deletions and domain swapping constructs have clearly identified transactivation abilities for the chick Sox1 protein (Kamachi et al. 1999). These studies have localized the transactivation domain to the C-terminal region, with two important subdomains corresponding to amino acid positions 135 – 255 (containing the PRD-type repeat) and 338 – 391 in the human SOX1 protein sequence. The same study presented strong evidence that Sox1 activation of crystallin expression is mediated by side-by-side binding with δEF3, an unidentified lens-specific cofactor. In studying the mechanisms of target selection by Sox 1, 2, 3, and 9 in the chick system, Kamachi and coworkers (1999) found that DNA-binding and transactivation activities of these Sox proteins were virtualy interchangeable. However, Sox9 could not substitute for Sox 1, 2, or 3 in lens-specific activation of δ-crystallin because of its inability to interact with δEF3 (Kamachi et al. 1999).

The authors hypothesized that the δEF3 partner is required to stabilize DNA-binding by the Sox1 HMG domain, and that δEF3 has lens-restricted expression. Similar stabilization by partner factors has been suggested for the HMG proteins LEF-1 (Carlsson et al. 1993) and Ste11 in yeast (Kjærulff et al. 1997). The functional domains identified are illustrated below in **Figure 36** (Kamachi et al. 1999), and the partner-mediated model for target binding is illustrated in **Figure 9**, in the Introduction (Kamachi et al. 2000).

Figure 36. Functional domains of Sox1, 2,and 3 proteins (Kamachi et al. 1999)



The transcriptional activation domain along with the known HMG DNA-binding domain and the nuclear localization signals earmark *SOX1* as a transcription factor. The studies in animal models emphasize a role for transcriptional regulation by *Sox1* in embryonic lens development. Chick Sox1, Sox2 and Sox3 proteins are members of a complex which binds the DC5 enhancer and activates lens-specific transcription of the δ1-crystallin gene (discussed above). In the developing murine eye, *Sox*2 and *Sox* 3 are highly expressed in the lens placode, with *Sox1* exhibiting subsequent high levels of expression in the lens germinal epithelium around the time of crystallin expression (see **Figure 6**, Introduction; Collignon et al. 1996; Kamachi et al. 1998). Homozygous *Sox1* deletion mice exhibit small eyes with cataracts, as well as spontaneous seizures. These deletion mice have severely reduced γ-crystallin gene expression which leads to impaired

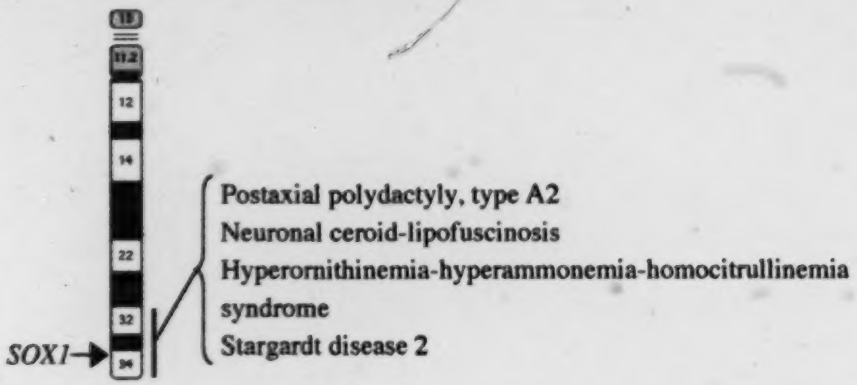
lens fibre elongation during development (see **Figure 7**, Introduction; Nishiguchi et al. 1998).

My research has not dissected the nature of *SOX1* expression in developing human neural or ocular tissues, but has laid the groundwork for such experiments by isolating and characterizing the human *SOX1* gene. Investigations into the human embryonic expression profiles for *SOX1*, *SOX2*, and *SOX3* could determine whether they exhibit similar patterns to those found in chick and mouse. In addition, the identification of proteins which interact with *SOX1*, *SOX2*, and *SOX3* would help to clarify the functional specificity of the *SOX* genes in lens development.

B. Localization inferences

My results mapped human *SOX1* to chromosome 13q34, in agreement with Malas et al. (1997) (**Figure 37**). This region of human chromosome 13 exhibits conserved synteny with mouse chromosome 8 (Koizumi et al. 1995), which correlates well with the mapping of murine *Sox1* to proximal chromosome 8 (Malas et al. 1996). No known developmental or ocular defects in the mouse have been mapped to this region of the mouse genome.

Figure 37. *SOX1* location on chromosome 13



The only ocular disorder that maps to the same region on chromosome 13 as *SOX1* is Stargardt disease 2, an autosomal dominant form of macular dystrophy with flecks (OMIM entry 153900). The recessive form of Stargardt disease maps to chromosome 1p, and results from mutations in the ATP-binding cassette transporter gene. However, a dominant form of Stargardt's was mapped to 13q34 by Zhang et al. (1994) using a large family from Indiana, and a few other families with dominant inheritance have been reported (Cibis et al. 1980). *SOX1* may be a candidate gene for Stargardt disease 2, since an autosomal dominant mode of inheritance has also been observed in *SOX9* and *SOX10* mutants. However, the evidence from the role of *SOX1* in mice and chick lens development and the expression profile I have observed do not suggest a role in retinal physiology. Furthermore, the results of radiation hybrid mapping do not place SOX1 within the critical region for Stargardt disease 2.

Although no particularly notable candidate human diseases have been associated with the *SOX1* locus, patients with deletions of the terminal region of the long arm of chromosome 13 exhibit facial dysmorphism, and one case had bilateral

aniridia (Brown et al. 1993; Rivera et al. 1984). **Table 11** depicts several 13q34 deletion cases reported in the literature; however, thepresence or absence of *SOX1* in these deletion patients is unknown. Recently, a chromosomal deletion map of human malformations was constructed based on information collected from 1,753 patients with nonmosaic contiguous deletions from The Human Cytogenetics Database (HCDB), a commercial catalog of postnatally ascertained, cytologically detectable chromosomal aberrations (Schinzel 1994). Based on analysis of this database, Brewer and coworkers found that deletion of the 13q34 band is strongly associated with microphthalmia, anal atresia, ambiguous genitalia, and holoprosencephaly (Brewer et al. 1998). Weaker associations for 13q34 include talipes equina varus, small bowel atresia, intestinal malrotation, coloboma, trigonocephaly, and microcephaly. Terminal deletions were not distinguished from interstitial deletions in this study. Human chromosomal deletions in general produce nonspecific phenotypic effects, such as growth retardation and developmental delay. However, loss of single gene effects are clearly detectable in some cases, such as the case of *PAX6*, which was initially identified because of its involvement in the WAGR syndrome (Wilms tumour, aniridia, genitourinary anomalies, and mental retardation) found among 11p13 deletion patients (Hanson and van Heyningen 1995). It is possible that *SOX1* deletions underlie the microphthalmia associated with the 13q34 region.

Table 11. Phenotypes of chromosome 13q34 deletion patients

	Brown et al. Case 8(1993)	Brown et al. Case 9 (1993)	Brown et al. Case 13(1993)	Brown et al. Case 14(1993)	Rivera et al. Case 1 (1984)	Rivera et al. Case 2 (1984)
Deletion	q32q34	q32q34	q22q34	q31.2q34	q32.3-qter(mat)	q32.3-qter
Age, sex	8y, M	4m, M	2m(died), M	Stillborn(38w), F	5m, F	2y7m, F
Size	normal	normal	2500g birth wt.	1200g birth wt.	2675g birth wt., delayed growth	2050g birth wt., delayed growth
Face/head	hypertelorism, facial asymm., depressed nasal bridge	hypertelorism	severe microcephaly, hypertelorism	large occipital encephalocele, severe microcephaly	low-set posterior hairline, facial asymm., hypertel., chubby cheeks	narrow forehead, chubby cheeks
Eyes		bilateral partial aniridia	fixed pupils		mongoloid palpebral fissures	left exotropia, epicanthic folds, mongoloid palpebral fissures, ptosis
Ears	low-set,	rotated			large lobules, prom. antitragus	large, abnormal
Nose		beaked			high nasal bridge	high nasal bridge
Mouth		micrognathia	high arch palate			protruding incisor
Brain		absence of corpus callosum			basal cisternae, large ventricles	
Develop.	psychomotor severe delay				developmental delay	developmental delay
Factor VII,X	heterozygous	50%				normal coagulation
Genital	Undes. testes	micropenis			normal	normal
Gut/Anal		displaced anus	imperfor. anus, recto-vaginal fistula, gut malrotation	anteriorly displaced anus		
Hands		normal fingers	absent thumbs	bilaterally absent thumbs	slender fingers	slender fingers, contractures, 5th finger clinodactyly
Feet		inturned toes	absent 2nd toes			broad first toes
Kidneys		normal	hypoplastic			
Heart			murmur, dilated aorta, ASD PDA		normal	normal
Muscle			hypotonia		hypotonia	hypotonia
Lung			abnormal lobes			
Bone					delayed bone age, abnor. ossification of vertebrate T3	prominent metopic suture

C. The *SOX1* transcript: several possibilities

1. Expression patterns

The *SOX1* expression profile which I observed is not easy to interpret. Essentially, I have detected two differently sized transcripts. The 4.6 kb transcript is expressed ubiquitously, with the highest levels in fetal kidney. In addition, this transcript is also present in undifferentiated NTera2 cells. This transcript was detected with several *SOX1* probes, including 5' UTR, HMG box-containing, and 3' ORF fragments. The 3.9 kb transcript was only detected in fetal brain. The two probes which detected this smaller transcript were both located downstream of the *SOX1* stop codon.

Other reports of *SOX1* expression focus on its specific presence in developing neural and ocular tissues. In the mouse, *Sox1*, *Sox*2, and *Sox* 3 are all expressed at high levels in the embryonic central nervous system. A 4.0 kb transcript was detected on a Northern made from whole embryo RNA, and *Sox1* expression was also detected in murine genital ridge and testis (Collignon et al. 1996; Wood and Episkopou 1999). Malas et al. (1997) report moderately abundant expression of a 3.9 kb *SOX1* transcript in 16 week human embryonic brain mRNA but not in NTera2 RNA. Multiple transcripts are not mentioned for *SOX1* in these other reports. However, the quality of the Northern photograph for the human *SOX1* report (Malas et al. 1997) is not high enough to make strong conclusions about the exact size or number of transcripts. I have not found evidence in the literature that *SOX1* expression has been assayed in other human tissues. My results do not preclude a role for *SOX1* in neural development, but the significant levels of *SOX1* expression in adult tissues, such as heart, liver, and pancreas, suggest *SOX1*

could have a broader role in maintenance of cell identity or control of transcription.

The relatively strong expression of human *SOX1* in fetal kidney suggests *SOX1* could function as a developmental regulator in renal tissue, and that a hereditary renal dysplasia could perhaps result from *SOX1* loss of function. Renal and ocular findings have been associated in diseases such as renal coloboma syndrome, which can be caused by mutations in *PAX2* (Sanyanusin et al. 1995), and various forms of Alport syndrome, which can be caused by type IV collagen mutations (Kashtan 1995). *SOX1* expression in human fetal kidney along with its eye-specific expression in other vertebrates suggest it may be involved in other renal and ocular pathologies.

2. Two genes?

I have considered several possibilities to explain the detection of two transcripts in the *SOX1* region. A simple interpretation would be that the two transcripts represent two different genes. The 4.6 kb transcript would likely be *SOX1*, since all probes within the *SOX1* ORF detected this transcript, and the 3.9 kb species would represent an exon from another, unknown gene. This hypothesis would account for the markedly distinct expression patterns observed for the two transcripts. It would also explain why the 3.9 kb transcript is not detected by most *SOX1* probes, including the HMG box domain, since this transcript would not include the *SOX1* ORF.

It would be tempting to assume that the two transcripts were situated adjacent to one another in the genomic sequence, with the end of the *SOX1* transcript falling

just before the second transcript began. If this breakpoint lay within the *Pst* I 1.75 kb fragment of the *SOX1* 3' region, it would explain the results of the 3' Northerns, in which the *Pst* I 1.75 kb fragment detected both transcripts, and subclones from either end of the *Pst* I 1.75 kb fragment detected only one transcript each (**Figure 25**). In support of this possibility, a small exon is predicted to lie within the 1.0 kb *Xho* I – *Pst* I subclone from the 3' end of the *Pst* I 1.75 kb fragment. Furthermore, a small gap exists in the EST and THC coverage of the *SOX1* 3'region just upstream of the predicted exon, and this gap falls just downstream of a predicted polyadenylation signal that would presumably terminate the *SOX1* transcript (**Figure 26**). Other examples of back to back genes have also been found in other areas of the genome, such as the *WT1* locus on 11p13.

Several observations cannot be accounted for well by the two adjacent gene theory. If the *SOX1* transcript does terminate in the gap preceding the small exon, at approximately position 2300 bp, at least 2 kb of additional, 5'UTR would have to be part of the transcript to make up the total 4.6 kb size. However, Northern probing with the series of 5' fragments suggested that the 5' end of the *SOX1* transcript falls between positions -800 and -1000 bp (**Figure 25**). An undetected 5' exon further upstream could account for the missing one kilobase of *SOX1* transcript. Although no exons in this 5' region were strongly predicted by the computer programs, an EST search of the region did reveal a 300 bp EST at position -1966 to -1665 bp. But perhaps the strongest argument against termination of the *SOX1* transcript around position 2300 bp is the numerous fetal brain cDNA clones I recovered that clearly span this region. The largest one, FB#8, begins in the coding region at position 682 bp and extends to position

2763bp. Although chimerism is relatively common in cDNA libraries, it is very unlikely that this clone would represent two different transcripts from the same region.

If the 3.9 kb transcript represents a second, distinct, gene, this gene may not be syntenic with *SOX1*. However, since the probe used to identify the chromosomal location of *SOX1* would have included the sequence that detected the 3.9 kb transcript by Northern, one would have expected another FISH signal (not at 13q34) indicating the location of the second gene. No such second signal was detected.

2. One gene, two transcripts?

A second explanation of the two transcripts is that they are alternate *SOX1* transcripts. Several other *SOX* genes have alternate transcripts, including *SOX2*, *SOX5*, *SOX6*, *SOX 11*, *SOX17*, and *SOX22* (Collignon et al., 1996; Lefebvre et al. 1998; Jay et al. 1995; Kania et al. 1996; Jay et al. 1997). The multiple transcripts can be explained by alternate splicing for *SOX5*, *SOX6*, and *SOX17*, but *SOX2*, *SOX11* and *SOX22* are monoexonic like *SOX1*, at least in the coding region. Perhaps untranslated exons are present in the upstream or downstream UTR domains of these so-called monoexonic genes, and these exons may play an important role in determining tissue specificity for each transcript.

There is significant evidence that the major 4.6 kb *SOX1* transcript starts at about position -1000 bp and extends to position 3500 bp. Evidence for the starting position comes from the Northern localization of the 5' end of the *SOX1* transcript, and there is a suitable transcription start site prediction at position -988

bp. Evidence for a termination site comes from the extent of EST and THC coverage of the *SOX1* 3' UTR, which is continuous until about position 3500 when considered together with the cDNA clones detected by my fetal brain library screening. This predicted *SOX1* genomic structure would be fully monoexonic, consistent with the lack of any strongly predicted exons in the nearby vicinity.

Although the 4.6 kb transcript can be accounted for, it is not easy to explain the 3.9 kb transcript as an alternative form of *SOX1*. The major difficulty is that probes within the *SOX1* ORF, even from the HMG box itself, do not pick up the 3.9 kb transcript on Northern blots. Also, there are not many alternative transcription start sites or polyadenylation sites predicted for the region, and there are no obvious gene feature combinations that would produce a 3.9 kb transcript. Another problem with this theory is that the 1.0 kb *Xho* I – *Pst* I probe from the 3' UTR should have detected the 4.6 kb transcript, but did not. The possibility remains that hidden intron/exon structures could still give rise to two *SOX1* transcripts with unidentified starting and terminating regions.

3. Other possibilities

The detection of the two transcript sizes may also be explained as artifactual. All of the probes used on Northern blots were at least 400 bp long and would not be expected to give unspecific hybridization. However, one of the transcripts (likely the 3.9 kb) may have been detected by weak hybridizaiton of a similar but non-identical sequence.

It is very unlikely that the 3' probes which detected the 3.9 kb transcript would cross-hybridize to another *SOX* gene, since *SOX* genes diverge widely outside of the HMG box region, and presumably even more widely outside of the coding region. Likewise the 5' probes which detected the 4.6 kb transcript would not be expected to bind another *SOX* gene. One alternative, though, is that one of the transcripts represents a *SOX1* pseudogene located elsewhere in the genome. The 3.9 kb transcript may be such a gene, since it appears to lack the HMG box domain. Although pseudogenes have been detected for other *SOX* genes, such as *SOX5*, BLAST searches do not show evidence of a pseudogene elsewhere in the genome for *SOX1*.

None of the alternative explanations proposed here can fully account for all of the data and *in silico* predictions about the *SOX1* genomic structure. I find the monoexonic structure for a 4.6 kb transcript to be quite convincing, especially since it relies primarily on experimental evidence, from the 5' UTR probes as well as the 3' cDNA coverage, rather than on *in silico* predictions. Instead of an alternative *SOX1* transcript, though, it seems more plausible that the 3.9 kb transcript represents a gene from elsewhere in the genome. Although a second locus was not detected by FISH, the sequence in the *SOX1* region with similarity to the second gene may not have been long enough to generate a strong FISH signal. Since the *SOX2* transcript size is reported at 3.5 kb, it is tempting to speculate that the 3.9 kb band is, in fact, *SOX2*. Once genomic sequence for the *SOX2* region is available, it will be possible to make direct sequence comparisons of the 3' regions of *SOX1* and *SOX2* to put this theory to the test.

D. A new cataract gene?

A mutation screen was carried out for the *SOX1* coding region. This screen included 18 patients from microphthalmia and cataract families, 21 patients with other anterior segment anomalies, and 63 control individuals. A total of four sequence alterations were detected in the *SOX1* gene (**Table 10**). Of these alterations, three were found among normal control individuals and one, the T1039C transition, was only detected in patient #927 and two other members of his family. Patient #927 was reported to have zonular cataracts, in which the opacification of the lens is confined to one layer, or zone, of the lens, usually the layer surrounding the fetal nucleus, which is clear (Ionides et al. 1999). The ocular phenotype of the brother to patient #927 is not known in detail. The son of patient #927, in whom the T1039C alteration was also found, had cataracts, microphthalmia and secondary glaucoma. Could these cataracts be due to a *SOX1* mutation? *SOX1* has a direct effect on γ-crystallin expression in mice, since *SOX1* deletion mice have severely reduced γ-crystallin levels (Nishiguchi et al. 1998), causing abnormally small lenses as well as cataracts. However, heterozygous deletion mice were not observed to have any detectable lens anomalies. The T1039C alteration found in patient #927 was a heterozygous change. Nonetheless, the effect of one altered *SOX1* allele may be more deleterious in humans than in mice. Recently, mutations in two of the γ-crystallin genes have been shown to underlie human cataracts (Héon et al. 1999, Ren et al. 2000, Stephan et al. 1999). It is certainly plausible that alterations in *SOX1*, a transcription factor that regulates γ-crystallin expression, could cause a similar phenotype.

The *SOX* gene mutations shown to cause human disorders to date are all dominant, and the detection of a heterozygous *SOX1* alteration in patient #927 would be consistent with a dominant mode of action. There is a possibility, however, that alteration of a single *SOX1* allele is not sufficient to cause a detectable phenotype. *Sox1, 2,* and *3* are all expressed in the developing mouse lens to varying degrees, and there may be functional redundancy among them that could help to compensate for loss of a single *Sox1* allele (Collignon et al. 1996). This compensation would explain the lack of observable anomalies in the *Sox1* heterozygous deletion mice. Alternatively, *Sox1* could be a modifier locus to another cataract pre-disposing gene. To test this possibility, *Sox1* deficiency could be studied on a variety of genetic backgrounds in mice.

If heterozygous *SOX1* mutations effect human cataracts, unlike the mouse, *SOX1* may have a somewhat different role in the human than the mouse eye. Furthermore, the *SOX1* transition we detected may dramatically modify *SOX1* transactivating abilities, rather than abrogating them, and have a dominant negative effect. Finally, the depth of clinical assessment of the human eye has not been equalled in murine physiology, and it may that there are, in fact, alterations in the heterozygous *Sox1* deletion mice that have not yet been identified.

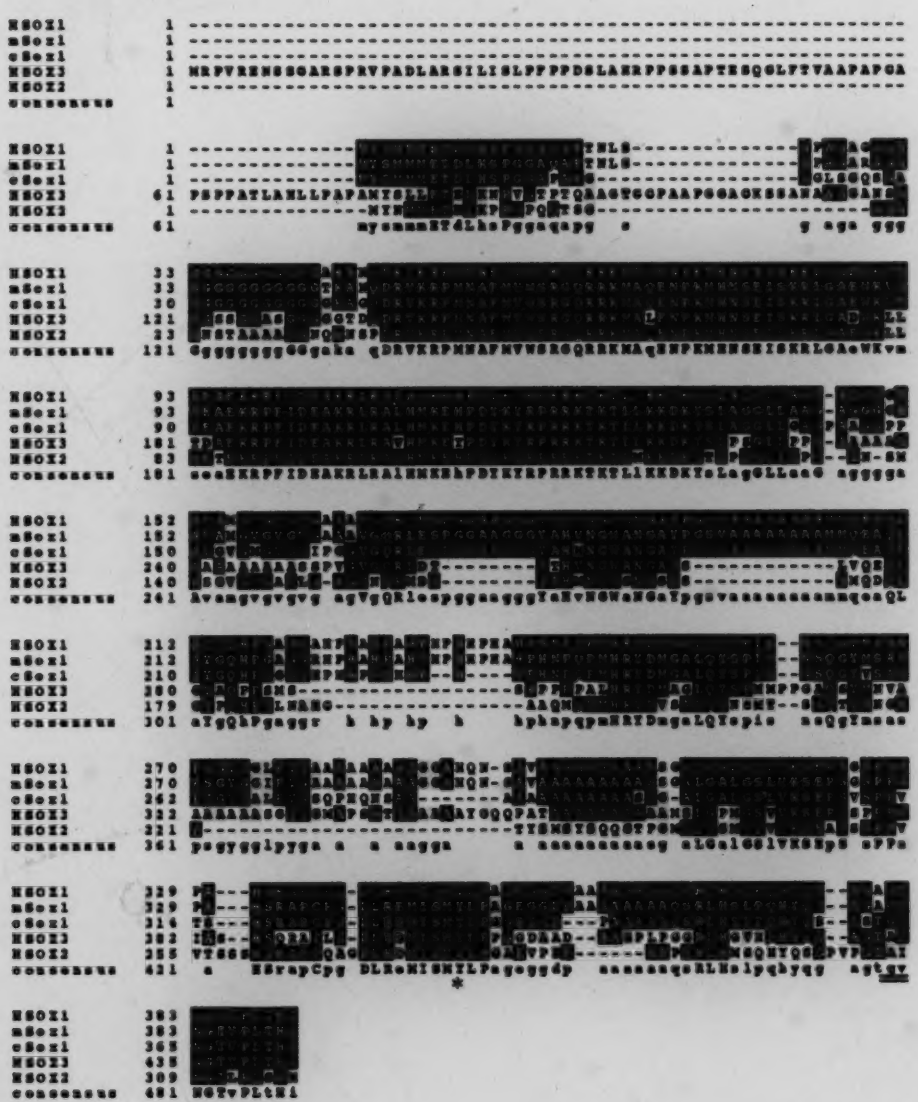
The T1039C change detected in patient #927 alters a tyrosine to a histidine at amino acid position 347 in the C-terminal region of the SOX1 protein. Activation studies of the DC5 enhancer by Sox1 in chick lens culture have identified two domains, at aa132 – 247 and aa 323 – 373, which are necessary for activation (Kamachi et al. 1999). The Y347H change in human SOX1 corresponds to aa 332 in chick Sox1, which is part of the distal activation domain. This residue is

found to be conserved among human, mouse, and chicken *SOX1*, as well as in the *SOX2* and *3* genes of all three species (**Figure 38**). The substitution of a histidine for a tyrosine is neither conservative nor overly dramatic, and speculation about the possible effect of such a change is difficult without further knowledge of the structure and function of that region of the SOX1 protein.

Final conclusions about the nature of the T1039C *SOX1* sequence change cannot be made until more clinical information is gathered about the other family members who carry the change. Ideally, additional cases of zonular cataracts should be screened for *SOX1* alterations, a process which has been initiated but not yet completed. It may be that *SOX1* heterozygotes have an increased risk of cataract development, with ultimate phenotypic outcome dependent on other genetic and environmental factors. Certainly one would predict that homozygotes for *SOX1* mutations would have a more severe phenotype than heterozygotes. The identification of homozygous null patients may or may not occur, depending on the requirement of *SOX1* for proper early embryonic neural development. Since *Sox1* deleted mice appear fully viable, it seems very likely that a similar outcome for humans may exist, in which case a fairly broad range of *SOX1* phenotypes may occur.

Figure 38. Comparative amino acid alignment of human, mouse, and chick *SOX* genes

Legend: Blackened regions are shared between at least three sequences; shaded regions are conservative changes. On the consensus sequence, SOX1 alterations detected in human subjects are indicated as follows: the position of the Y347H substitution is marked by asterisk, and the duplicated GV residues at position 381 are double underlined.



E. Campomelic dysplasia mutation

The ins1456G mutation that we identified is the first frameshift mutation reported, to date, in a CD patient that has survived the first year of life. This frameshift is predicted to affect the SOX9 sequence beginning at amino acid 485, leaving the HMG domain intact but altering approximately half of the region needed for full transactivating domain function. The distal 107 amino acids of SOX9 are responsible for the transactivating ability of the protein, and amino acids 466-509 are necessary for optimal activity (Südbeck et al. 1996). Biochemical studies by McDowall et al. (1999) have recently shown that C-terminal deletions of *SOX9* cause loss of transactivation activity. In their study, a SOX9 deletion construct containing only the first 485 amino acids of the protein had significantly reduced transactivation ability in comparison with wildtype SOX9, but did retain appreciable activity. This 485 amino acid deletion construct is missing almost exactly the region that is altered in our patient by the frameshift shown in **Figure 39**. Similarly, a SOX9 construct truncated even more, at codon 440, also showed low level activation, and three patients with this truncation at codon 440 survived infancy (Meyer et al. 1997). These studies suggest that the CD patient we studied may have survived infancy due to residual SOX9 activity. However, another patient with a 4 bp insertion at codon 507, leaving virtually all of the transactivation domain intact, died neonatally (Kwok et al. 1995).

Strict predictions of phenotype based on *SOX9* genotype are clearly impossible. An overview of *SOX9* mutations in cases of CD is shown in **Figure 39**. Frameshift and truncation mutations appear to be the most deleterious as few individuals carrying these mutations have survived early infancy. This observation is consistent with *SOX9* haploinsufficiency as the underlying cause of

CD, since frameshift and truncation mutations lead to the most severe predicted reduction of SOX9 protein function. Two mutations, ins1104A (Kwok et al. 1995) and C1692G (Wagner et al. 1994; Meyer et al. 1997; Hageman et al. 1998), have been observed several times in unrelated individuals, potentially indicating functionally important residues or mutational hotspots. There is no correlation between mutation and sex reversal (Meyer et al. 1997).

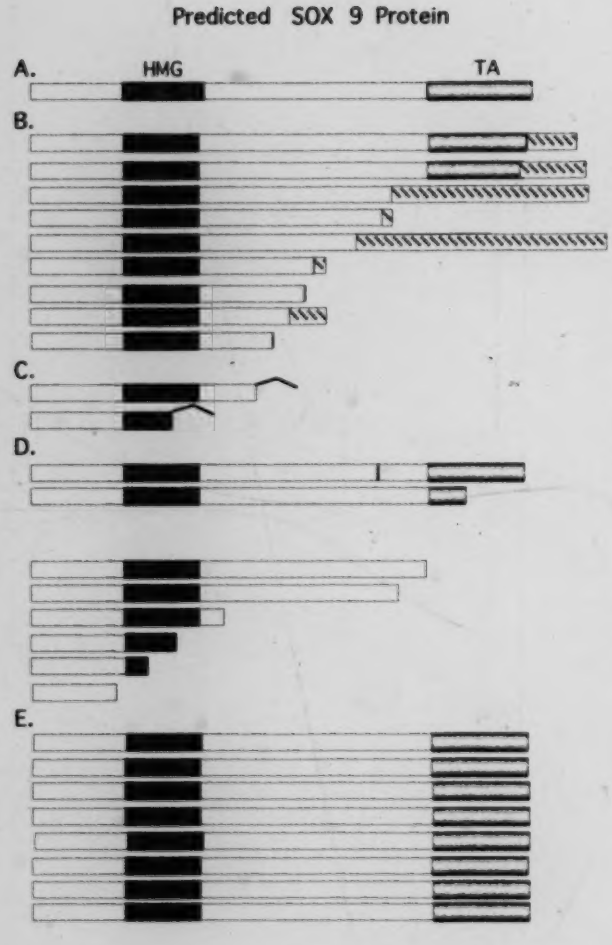
The features of *SOX9* mutations in campomelic dysplasia are characteristic of transcription factor mutations in general. First, these genes show a high degree of dosage sensitivity, and inactivating mutations (truncations and frameshifts) tend to be most severe. The degree to which the mutant SOX9 protein from our patient retained residual activity may have enhanced her survival time. Secondly, phenotypes are variable. Thus other patients, with seemingly less severe mutations, did not survive infancy. Finally, the effects of transcription factor mutations tend to be widespread, consistent with the effects of *SOX9* mutations on sexual development, chondrogenesis, as well as general developmental processes.

Figure 39. Summary of *SOX9* mutations in campomelic dysplasia.

Legend: **A**. Normal SOX9 protein with DNA-binding (HMG) and transactivation (TA) domains. **B**. Frameshift mutations **C**. Splicing mutations **D**. Nonsense and deletion mutations **E**. Missense mutations; substitution at asterisk (*).

Survival: neonatal death (-); prior to 1 year (-/+); beyond 1 year (+); not indicated (NI).

** Mutation G827C was identified in an acampomelic dysplasia patient (lacking bowing of the long bones). Abbreviations: M (male), F (female), ins (insertion), del (deletion), fs (frameshift), pt (protein truncation), sp (splice site), int (intron), nucleotide (nt), amino acid (aa)



Karyotype	Mutation Nt Position	Mutation Aa Change	Survival	Reference
46, XX, F	ins 1519 4bp	fs507	-	Kwok et al. 1995
46, XX, F	ins 1456 G	fs485	+	Present case
46, XY, M; 46, XY, F	ins 1103 A	fs368	-, -	Kwok et al. 1995
46, XY, F	del 1070 43bp	fs357	-	Meyer et al. 1997
46, XY, F	ins 985 G	fs329	+/-	Wagner et al. 1994
46, XY, F	ins 858 4bp	fs287	-	Foster et al. 1994
46, XX, F	del 819 10bp	fs277	-	Meyer et al. 1997
46, XY, F	ins 789 G	fs263	NI	Foster et al. 1994
46, XY, F	ins 736 C	fs246	-	Cameron et al. 1996
46, XY, F	5' sp, int 2	pt	-	Wagner et al. 1994
46, XY, F	3' sp, int 1	pt	-	Kwok et al. 1995
46, XY, F	del 1061 9bp	del 354	+/-	Wagner et al. 1994
46, XY, M; 46 XX, F; 46 XY, F	C1320G	Y440X	+, +, +	Hageman et al. 1998; Meyer et al. 1997; Wagner et al. 1994
46, XY, M	G1198T	E400X	-	Meyer et al. 1997
46, XY, F	C1123T	Q375X	-	Meyer et al. 1997
46, XX, F	C583T	Q195X	NI	Foster et al. 1994
46, XY, F	G442T	E148X	+/-	Wagner et al. 1994
46, XY, F	C349T	Q117X	+	Meyer et al. 1997
46, XX, F	G257A	W86X	-	Meyer et al. 1997
46, XY, F	C323T	P108L	-/+	Meyer et al. 1997
46, XY, M	T334C	F112L	-	Kwok et al. 1995
46, XY, M	T335C	F112S	+	Goji et al. 1998
46, XX, F	C356T	A119V	-	Kwok et al. 1995
46, XY, F	T427C	W143R	-/+	Meyer et al. 1997
46, XX, F	G455C**	R152P	-	Meyer et al. 1997
46, XY, M	T480C	H165Y	+	McDowall et al. 1999
46, XY, inv(9), M	C509G	P170R	-	Meyer et al. 1997

CONCLUSIONS

My research has helped to explore the structure and putative functions of three human *SOX* genes, *SOX1*, *SOX2* and *SOX9*. My study of *SOX2* focused primarily on its expression pattern, which was focused in human fetal and adult brain. For *SOX9*, I took part in a mutation search for a campomelic dysplasia patient, and we identified a guanine insertion at nucleotide 1456 which is predicted to disrupt the terminal 24 amino acids of the this patient's *SOX9* gene. By surveying the effects of other *SOX9* mutations and looking at other biochemical studies of the gene, we formulated the proposal that the retention of residual transactivation ability in the SOX9 mutant protein may explain why our patient survived past infancy.

The most intensive focus of my work was the human *SOX1* gene. I sequenced *SOX1* from genomic DNA, discovered its chromosomal location, profiled its expressed pattern in human tissues, and conducted a *SOX1* mutation screen on patients with anterior segment anomalies. Although the 76%, GC content presented some challenges to sequencing *SOX1*, my results suggest that *SOX1* has a 1173 bp open reading frame, and codes for a 391 amino acid protein. The human DNA and protein sequences I determined for *SOX1* are 95% and 99% identical to the murine *Sox1* sequences, respectively. Both fluorescence *in situ* hybridization and radiation hybrid mapping allowed me to localize *SOX1* to human chromosome 13q34. At this time there are not any obvious human disorders mapped to this region of chromosome 13 which may be caused by *SOX1* mutations.

I detected widespread expression of a 4.6 kb transcript using *SOX1*-specific probes on multiple tissue human fetal and adult Northern blots. Using a probe

just downstream of the *SOX1* open reading frame, I also found evidence for a 3.9 kb transcript expressed only in fetal brain. DNA samples from patients with microphthalmia, cataracts and other anterior segment disorders were tested for *SOX1* mutations by single stranded conformation polymorhpism analysis and sequencing. I detected a thymine to cytosine transition at position 1039 in *SOX1* that was not present among 126 normal chromosomes tested. This alteration was found in a patient with zonular cataracts and two of his family members. No disease-causing mutations have previously been reported for *SOX1*, so this alteration may represent a new disease-causing mutation in a *SOX* gene. However, additional evidence of similar changes in other cataract patients is still required before drawing conclusions.

Taken together, my results support the idea that *SOX1* is a highly conserved gene which likely has a significant function in human development. I feel that *SOX1*, like *SOX9*, may play multiple roles in the complex story of embryonic morphogenesis. The broad expression profile observed for *SOX1* by Northern blots is one piece of evidence for a diversity of functions. There may be tissue-specific alternate transcripts for *SOX1*, which would also suggest different activities in different tissues. Studies in animal models affirm a role for *SOX1* in lens development which is mediated by a partner factor, δEF3. The ability to interact with diverse partners may multiply the roles for *SOX1* as well, especially since the *SOX* target sequence is relatively small and common in the genome. *SOX1* itself may be the target of repression by the *WT1* gene, and, since it is expressed strongly in fetal kidney, perhaps affect early renal development.

I have established a basic groundwork for examining *SOX1* in the context of human development, and future experiments will undoubtedly reveal further details of its structure and function. Perhaps most inviting is the incentive to expand the *SOX1* mutation screen to find other cataract patients who carry *SOX1* mutations, either as heterozygotes or possibly as homozygous null alleles, which may have a more dramatic phenotype. Another pressing question is the identity of partner factors for *SOX1*, both in the developing lens as well as in other tissues. I also look forward to additional data that will help define the *SOX1* transcript, and to characterize its promoter region. It will be as interesting to see what is controlling *SOX1* as it is to see what *SOX1* controls. The journey of discovery for *SOX1* has just begun, and it promises to be a rich one in the context of the ongoing explosion of embryonic, genomic, proteomic, and bioinformatic knowledge that confronts us.

(So long, and thanks for all the FISH!)

REFERENCES

Abdelhak S, Kalatzis V, Heilig R, Compain S, Samson D, Vincent C, Weil D, Cruaud C, Sahly I, Leibovici M, Bitner-Glindzicz M, Francis M, Lacombe D, Vigneron J, CharachonR, Boven K, Bedbeder P, Van Regemorter N, Weissenbach J, Petit C (1997) A human homologue of the Drosophila eyes absent gene underlies branchio-oto-renal (BOR) syndrome and identifies a novel gene family. *Nat Genet* 15: 157-164.

Amiel J, Salomon R, Jolly D, Delezoide, A-L, Auge J, Gubler M-C, Munnich A, Lyonnet S, Antignac C, Vekemans M, Broyer M, Attie-Bitach T (1998) PAX2 expression during early human development and its mutations in renal hypoplasia with or without coloboma. (Abstract) *Am J Hum Genet* 63 (suppl.): A7.

Andrews PW, Damjanov I, Simon D, Banting GS, Carlin C, Dracopoli NC, Føgh J (1984) Pluripotent embryonal carcinoma clones derived from the human teratocarcinoma cell line Tera-2: differentiation *in vivo* and *in vitro*. *Lab Invest* 50: 147-162

Arsic N, Rajic T, Stanojcic S, Goodfellow PN, Stevanovic M (1998) Characterization and mapping of the human *SOX14* gene. *Cytogenet Cell Genet* 83: 139-146.

Azuma N, Hirakiyama A, Inoue T, Asaka A, Yamada M (2000) Mutations of a human homologue of the Drosophila eyes absent gene (*EYA1*) detected in patients with congenital cataracts and ocular anterior segment anomalies. *Hum Molec Genet* 9: 363-366.

Azuma N, Nishina S, Yanagisawa H, Okuyama T, Yamada M (1996) *PAX6* missense mutation in isolated foveal hypoplasia.(Letter) *Nat Genet* 13: 141-142.

Azuma T, Seki N, Yoshikawa T, Saito T, Masuho Y, Muramatsu M (2000) cDNA cloning, tissue expression, and chromosome mapping of human homolog of *SOX18*. *J Hum Genet* 45: 192-195.

Baldwin CT, Hoth CF, Amos JA, da-Silva EO, Milunsky A (1992) An exonic mutation in the HuP2 paired domain gene causes Waardenburg's syndrome. *Nature* 355: 637-638.

Baldwin CT, Hoth CF, Macina RA, Milunsky A (1995) Mutations in *PAX3* that cause Waardenburg syndrome type I: ten new mutations and review of the literature. *Am J Med Genet* 58: 115-122.

Barishak Y (1992) Embryology of the eye and its adnexae. [Review] [187 refs] *Developments in Ophthalmology*. 24: 1-142.

Behlke MA, Bogan JS, Beer-Romero P, Page DC (1993) Evidence that the SRY protein is encoded by a single exon on the human Y chromosome. *Genomics* 17: 736-739.

Bell DM, Leung KK, Wheatley SC, Ng LJ, Zhou S, Ling KW, Sham MH, Koopman P, Tam PP, Cheah KS (1997) SOX 9 directly regulates the type-II collagen gene. *Nat Genet* 16: 174-178.

Berry V, Francis P, Kaushal S, Moore A, Bhattacharya S (2000) Missense mutations in MIP underlie autosomal dominant 'polymorphic' and lamellar cataracts linked to 12q. *Nat Genet* 25: 15-17.

Berta P, Hawkins JR, Sinclair A, Taylor A, Griffiths BL, Goodfellow PN, Fellous M, (1990) Genetic evidence equating SRY and the male sex determining gene. *Nature* 348: 248-250.

Bianchi ME, Beltrame M (1998) Flexing DNA: HMG-box proteins and their partners. *Am J Hum Genet* 63: 1573-1577.

Bondurand N, Kobetz A, Pingault V, Lemort N, Encha-Razavi F, Couly G, Goerich DE, Wegner M, Abitbol M, Goossens M (1998) Expression of the *SOX10* gene during human development. *FEBS Lett* 432: 168-172.

Bondurand N, Kuhlbrodt K, Pingault V, Enderich J, Sajus M, Tommerup N, Warburg M, Hennekam RCM, Read,AP, Wegner M, Goossens M (1999) A molecular analysis of the Yemenite deaf-blind hypopigmentation

syndrome: SOX10 dysfunction causes different neurocristopathies. *Hum Molec Genet* 8: 1785-1789.

Bopp D, Burri M, Baumgartne S, Frigerio G, Noll M (1986) Conservation of a large protein domain in the segmentation gene paired and in functionally related genes of Drosophila. *Cell* 47: 1033-1040.

Botquin V, Hess H, Fuhrmann G, Anastassiadis C, Gross MK, Vriend G, Schöler HR (1998) New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2. *Genes Dev* 12: 2073-2090.

Brewer C, Holloway S, Zawalnyski P, Schinzel A, FitzPatrick D (1998) Human Deletion Map. *Am J Hum Genet* 63: 1153-1159.

Bricarelli FD, Fraccaro M, Lindsten J, Muller U, Baggio P, Carbone LD, Hjerpe A, Lindgren F, Mayerova A, Ringertz H, Ritzen EM, Rovetta DC, Sicchero C, Wolf U (1981) Sex-reversed XY females with campomelic dysplasia are H-Y negative. *Hum Genet* 57: 15-22.

Brown S, Gersen S, Anyane-Yeboa K, Warburton D (1993) Preliminary definition of a "critical region" of chromosome 13 in q32: Report of 14 cases with 13q deletions and review of the literature. *Am J Med Genet* 45: 52-59.

Burmeister M, Novak J, Liang M-Y, Basu S, Ploder L, Hawes NL, Vidgen D, Hoover F, Goldman D, Kalnins VI, Roderick TH, Taylor BA, Hankin MH, McInnes RR (1996) Ocular retardation mouse caused by Chx10 homeobox null allele: impaired retinal progenitor proliferation and bipolar cell differentiation. *Nat Genet* 12: 376-384.

Busch SJ, Sassone CP, (1990) Dimers, leucine zippers and DNA-binding domains. *Trends Genet* 6: 36-40.

Cai J, Lan Y, Appel L, Weir M (1994) Dissection of the Drosophila paired protein: Functional requirements for conserved motifs. *Mech Dev* 47: 139-150.

Cameron FJ, Hageman RM, Cooke-Yarborough C, Kwok C, Goodwin LL, Sillence DO, Sinclair AH (1996) A novel germ line mutation in SOX9 causes familial campomelic dysplasia and sex reversal. *Hum Mol Genet* 5: 1625-1630.

Capel B, Swain A, Nicolis S, Hacker A, Walter MA, Koopman P, Goodfellow PN, Lovell-Badge R (1993) Circular transcripts of the testis determining geneSry in the adult mouse testis. *Cell* 73: 1019-1030.

Carlsson P, Waterman ML, Jones KA (1993) The hLEF-1/TCF-1 HMG protein contains a context-dependent transcriptional activation domain that induces the TCR enhancer in T cells. *Genes Dev* 7: 2418-2430.

Catron KM, Zhang H, Marshall SC, Inostroza JA, Wilson JM, Abate C (1995) Transcriptional repression by Msx-1 does not require homeodomain DNA-binding sites. *Mol Cell Biol* 15: 861-871.

Chalepakis G, Tremblay P, Gruss P (1992) Pax genes, mutants and molecular function. *J Cell Sci* 16: 61-67.

Chardard D, Chesnel A, Gozé C, Dournon C, Berta P (1993) Pw Sox-1: the first member of the Sox gene family in Urodeles. *Nuc Acids Res* 21: 3576.

Church GM, Gilbert W (1984) Genomic sequencing. *Proc Natl Acad Sci USA* 81: 1991-1995.

Cibis GW, Morey M, Harris DJ (1980) Dominantly inherited macular dystrophy with flecks (Stargardt). *Arch Ophthal* 98: 1785-1789.

Clépet C, Schafer AJ, Sinclair AH, Palmer MS, Lovell-Badge R, Goodfellow PN (1993) The human SRY transcript. *Hum Mol Genet* 2: 2007-2012.

Collignon J, Shanthini S, Hacker A, Cohentannoudji M, Norris D., Rastan S, Stevanovic M, Goodfellow PN, Lovell-Badge R (1996) A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. *Development* 122: 509-520.

Connor F, Cary PD, Read CM, et al. (1994) DNA binding and bending properties of the post-meiotically expressed Sry-related protein Sox-5. *Nuc Acids Res* 22: 3339-3346.

Connor F, Wright E, Denny P, Koopman P, Ashworth A (1995) The Sry-related HMG box-containing gene Sox6 is expressed in the adult testis and developing nervous system of the mouse. *Nuc Acids Res* 23: 3365-3372.

Coriat AM, Muller U, Harry JL, Uwanogho D, Sharpe PT (1993) PCR Amplification of *Sry*-related sequences reveals evolutionary conservation of the SRY-box motif. *PCR Methods and Applications* 2: 218-222.

Corliss CE (1976) *Patten's human embryology: elements of clinical development.* Toronto: McGraw-Hill.

Czerny T, Halder G, Kloter U, Souabni A, Gehring W, Busslinger M (1999) Twin of eyeless, a second Pax 6 gene of Drosophila, acts upstream of eyelessin the control of eye development. *Molecular Cell* 3: 297–307.

De Chen J, Bapat B, Bascom R, Willard H, Gallie B, McInnes RR (1989) Identification of a developmentally regulated human retinal homeobox gene. (Abstract) *Am J Hum Genet* 45: A111.

De Santa Barbara P, Bonneaud N, Boizet B, Desclozeaux M, Moniot B, Sudbeck P, Scherer G, Poulat F, Berta P (1998) Direct interaction of SRY-related protein SOX 9 and steroidogenic factor 1 regulates transcription of the human anti-Müllerian hormone gene. *Mol Cell Biol* 18: 6653-6665.

Denny P, Swift S, Brand N, Dabhade N, Barton P, Ashworth A (1992) A conserved family of genes related to the testis determining gene, SRY. *Nuc Acids Res* 20: 2887.

Denny P, Swift S, Connor F, Ashworth A (1992) An SRY-related gene expressed during spermatogenesis in the mouse encodes a sequence-specific DNA-binding protein. *EMBO Journal* 11: 3705-3712.

Desplan C (1997) Eye development: Governed by a dictator or a junta? *Cell* 91: 861-864.

Doward,W, Perveen R, Lloyd IC, Ridgway AEA, Wilson L, Black GCM (1999) A mutation in the RIEG1 gene associated with Peters' anomaly. *J Med Genet* 36: 152-155.

Duncan MK, Haynes JI II, Cvecl A, Piatigorsky J (1998) Dual roles for Pax-6: a transcriptional repressor of lens fiber cell-specific -crystallin genes. *Mol Cell Biol* 18: 5579-5586.

Dunn TL, Mynett-Johnson L, Wright EM, Hosking BM, Koopman PA, Muscat GEO (1995) Sequence and expression of Sox-18 encoding a new HMG-box transcription factor. *Gene* 161: 223-225.

Ebensperger C, Jager RJ, Latterman U, Dagna Bricarelli F, Keutel J, Lindsten J, Rehder H, Muller U, Wolf U (1991) No evidence of mutations in four candidate genes for male sex determination/differentiation in sex-reversed XY females with campomelic dysplasia. *Annales de Genetique* 34: 233-238.

Ekker SC, Ungar AR, Greenstein P, Von Kessler DP, Porter JA, Moon RT, Beachy PA (1995) Patterning activity of vertebrate hedgehog proteins in the development eye and brain. *Curr Biol* 5: 944-955.

Emanuel B, Zakai E, Moreau L, Coates P, Orrechio E. (1979) Interstitial deletion 13q33 resulting from maternal insertional translocation. *Clin Genet* 16: 340-346.

Engelkamp D, van Heyningen V (1996) Transcription factors in disease. *Curr Opin Genet Dev* 6: 334-342.

Epstein DJ, Vekemans M, Gros P (1991) Splotch (Sp-2H), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3. *Cell* 67: 767-774.

Fan YS, Davies LM, Shows TB (1990) Mapping small DNA sequences by fluorescent in situ hybridization directly on banded metaphase chromosomes. *Proc Natl Acad Sci USA* 87: 6223-6227.

Farr CJ, Easty DJ, Ragoussis J, Collignon J, Lovell-Badge R, Goodfellow PN (1993) Characterization and mapping of the human SOX4 gene. *Mamm Genome* 4: 577-584.

Ferrari S, Harley VR, Pontigiggia A, Goodfellow PN, Lovell-Badge R, Bianchi ME (1992) SRY, like HMG 1, recognizes sharp angles in DNA. *EMBO J* 11: 4497-4506.

Fini ME, Strissel KJ, West-Mays JA (1997) Perspectives on eye development. *Dev Genet* 20: 175-185.

Foster J, Dominguez-Steglich M, Guioli S, Kowk G, Weller PA, Stevanovic M, Weissenbach J, Mansour S, Young ID, Goodfellow PN, Brook JD, Schafer A (1994) Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature* 372: 525-530.

Foster JW, Graves JA (1994) An SRY-related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis determining gene. *Proc Nat Acad Sci USA* 91: 1927-1931.

Francis PJ, Berry V, Moore AT, Bhattacharya S (1999) Lens biology: development and human cataractogenesis [Review] *Trends Genet* 15: 191-196.

Freund CL, Gregory-Evans CY, Furukawa T, Papaioannou M, Looser J, Ploder L, Bellingham J, Ng D, Herbrick JA, Duncan A, Scherer SW, Tsui LC, Loutradis-Anagnostou A, Jacobson SG, Cepko CL, Bhattacharya,SS, McInnes RR (1997) Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. *Cell* 91: 543-553.

Freund CL, Horsford DJ, McInnes RR (1996) Transcription factor genes and the developing eye: a genetic perspective. [Review] [234 refs] *Hum Mol Genet* 5 Spec No: 1471-1488.

Freund CL, Wang Q-L, Chen S, Muskat BL, Wiles CD, Sheffield VC, Jacobson SG, McInnes RR, Zack DJ, Stone EM (1998) De novo mutations in the

CRX homeobox gene associated with Leber congenital amaurosis. (Letter) *Nat Genet* 18: 311-312.

Friedman JS, Walter MA (1999) Glaucoma genetics, present and future. [Review] [62 refs] *Clin Genet* 55: 71-79.

Furukawa T, Morrow EM, Cepko CL (1997) Crx, a novel otx-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. *Cell* 91: 531-541.

Gage PJ, Suh H, Camper SA (1999) Dosage requirement of Pitx2 for development of multiple organs. *Development* 126: 4643-4651.

Gehring W, Ikeo K(1999) Mastering eye morphogenesis and evolution. *Trends Genet* 15: 371-377.

Giese K, Cox J, Grosschedl R (1992) The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell* 69: 1-20.

Glaser T, Jepeal L, Edwards JG, Young SR, Favor J, Maas RL (1994) PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. *Nat Genet* 7: 463-471.

Goji K, Nishijima E, Tsugawa C, Nishio H, Pokharel RK, Matsuo M (1998) Novel missense mutation in the HMG box of the SOX 9 gene in a Japanese XY male resulted in campomelic dysplasia and severe defect in masculization. *Hum Mutat Suppl* 1: S114-S116.

Goze C, Poulat F, Berta P (1993) Partial cloning of SOX-11 and SOX-12, two new human SOX genes. *Nuc Acids Res* 21: 2943.

Graves JAM (1998) Interactions between *SRY* and *SOX* genes in mammalian sex determination. *BioEssays* 20: 264-269.

Graw J (1997) The crystallins: genes, proteins and diseases. *Biol Chem* 378: 1331-1348.

Griffiths R (1991) The isolation of conserved DNA sequences related to the human sex-determining region Y gene from the lesser black-backed gull (*Larus fuscus*). *Proc Royal Soc London* B 244: 123-128.

Grosschedl R, Giese K, Pagel J (1994) HMG domain proteins: architectural elements in the assembely of nucleoprotein structures. *Trends Genet* 10: 94-100.

Gubbay J, Collignon J, Koopman P, Capel B, Economou A, Münsterberg A, Vivian N, Goodfellow P, Lovell-Badge R (1990) A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* 346: 245-250.

Hacker A, Capel B, Goodfellow PN, Lovell-Badge R (1995) Expression of Sry, the mouse sex determining gene. *Development* 121: 1603-1614.

Hageman R, Cameron F, Sinclair A (1998) Mutation Analysis of the SOX 9 gene in a patient with campomelic dysplasia. *Hum Mutat Suppl* 1: S112-S113.

Hagiwara M, Klewer SE, Samson RA, Erickson DT, Lyon MF, Brilliant MH (2000) Sox6 is a candidate gene for p^{100H} myopathy, heart block, and sudden neonatal death. *Proc Natl Acad Sci USA* 97: 4180-4185.

Halder G, Callaerts P, Gehring W (1995) Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science* 27: 1788-1792.

Halder G, Callaerts P, Gehring W (1995) New perspectives on eye evolution. *Curr Opin Genet Dev* 5: 602-609.

Han K, Manley JL. (1993) Functional domains of the Drosophila Engrailed protein. *EMBO J* 12: 2723-2733.

Hanley NA, Hagan DM, Clement-Jones M, Ball SG, StrachanT. Salas-Cortés L, McElreavey, Lindsay S, Robson S, Bullen P, Ostrer H, Wilson DI (2000) SRY, SOX9, and DAX1 expression patterns during human sex determination and gonadal development. *Mech Dev* 91: 403-407.

Hanna-Rose W, Han,M (1999) COG-2, a Sox domain protein necessary for establishing a functional vulval-uterine connection in Caenorhabditis elegans. *Development* 126: 169-179.

Hanson IM, Churchill A, Love J, Axton R, Moore T, Clarke M, Meire F, van Heyningen V (1999) Missense mutations in the most ancient residues of the PAX6 paired domain underlie a spectrum of human congenital eye malformations. *Hum Mol Genet* 8: 165-172.

Hanson IM, Fletcher JM, Jordan T, et al. (1994) Mutations at the PAX6 locus are found in heterogeneous anterior segment malformations including Peters' anomaly. *Nat Genet* 6: 168–173.

Hanson I, van Heyningen V (1995) Pax6: more than meets the eye. *Trends Genet* 11: 268-272.

Haqq CM, King CY, Ukiyama E, Falsafi S, Haqq T, Donahoe P, Weiss M (1994) Molecular basis of mammalian sexual determination: Activation of Müllerian Inhibiting Substance gene expression by SRY. *Science* 266: 1494-1500.

Hargrave M, James K, Nield K, Toomes C, Georgas K, Sullivan T, Verzijl H, Oley C, Little M, De Jonghe P, Kwon J, Kremer H, Dixon M, Timmerman V, Yamada T, Koopman P (2000) Fine mapping of the neurally expressed gene *SOX14* to human 3q23, relative to three congenital diseases. *Hum Genet* 106: 432-439.

Hargrave M, James K, Yamada T, Koopman P (2000) *Sox14* maps to mouse chromosome 9 and shows no mutations in the neurological mouse mutants *ducky* and *tippy*. *Mamm Genome* 11: 231-233.

Hargrave M, Karunaratne A, Cox L, Wood S, Koopman P, Yamada T (2000) The HMG box transcription factor gene Sox14 marks a novel subset of ventral interneurons and is regulated by sonic hedgehog. *Dev Biol* 219: 142-53.

Hargrave M, Wright E, Kun J, Emery J, Cooper L, Koopman P (1997) Expression of the Sox11 gene in mouse embryos suggests roles in

neuronal maturation and epithelio-mesenchymal induction. *Dev Dyn* 210: 79-86.

Harley VR, Lovell-Badge R, Goodfellow PN (1994) Definition of a consensus DNA-binding site for SRY. *Nuc Acids Res* 22: 1500-1501.

Hejtmancik JF (1998) The genetics of cataract: Our vision becomes clearer. *Am J Hum Genet* 62: 520-525.

Hejtmancik JF, Kaiser MI, Piatigorsky J (1995) "Molecular biology and inherited disorders of the eye lens." Chapter 146 in *The Metabolic and Molecular Bases of Inherited Disease*, Scriver CR, Beaudet AL, Sly WS, Valle D (eds) Toronto: McGraw-Hill, p.4325-4349, 1995.

Heon E, Priston M, Schorderet DF, Billingsley GD, Girard PO, Lubsen N, Munier FL (1999) The gamma-crystallins and human cataracts: a puzzle made clearer. *Am J Hum Genet* 65: 1261-1267.

Herbarth B, Pingault V, Bondurand N, Kuhlbrodt K, Hermans-Borgmeyer I, Puliti A, Lemort N, Goossens M, Wegner M (1998) Mutation of the Sry-related Sox10 gene in Dominant megacolon, a mouse model for human Hirschsprung disease. *Proc Natl Acad Sci USA* 95: 5161-5165.

Hill RE, Favor J, Hogan BLM, Ton CCT, Saunders GF, Hanson IM, Prosser J, Hastie ND, Van Heyningen V (1991) Mouse small eye results from mutations in a paired-like homeobox-containing gene. *Nature* 354: 522-525.

Hill RE, Hanson IM (1992) Molecular genetics of the Pax gene family. *Curr Opin Cell Biol* 4: 967-972.

Hiraoka Y, Ogawa M, Sakai Y, Taniguchi K, Fujii T, Umezawa A, Hata J-I, Aiso S (1998) Isolation and expression of a human SRY-related cDNA, hSOX20. *Biochim Biophys Acta* 1396: 132-137.

Holst BD, Wang Y, Jones FS, Edelman GM (1997) A binding site for Pax proteins regulates expression of the gene for the neural cell adhesion

molecule in the embryonic spinal cord. *Proc Natl Acad Sci* USA 94: 1465–1470.

Houston C, Opitz J, Spranger J, Macpherson RI, Reed MH, Gilbert EF, Hermann J, Schinzel A (1983) The campomelic syndrome: review, report of 17 cases, and follow up on the currently 17-year-old boy first reported by Maroteaux et al. in 1971. *Am J Med Genet* 15: 3-28.

Ionides A, Francis P, Berry V, Mackay D, Bhattacharya S., Shiels A, Moore A (1999) Clinical and genetic heterogeneity in autosomal dominant cataract. *Br J Ophthalmol* 83: 802–808.

Irvine AD, Corden LD, Swensson O, Swensson B, Moore JE, Frazer DG, Smith FJD, Knowlton RG, Christophers E, Rochels R, Uitto J, McLean WHI (1997) Mutations in cornea-specific keratin K3 or K12 genes cause Meesmann's corneal dystrophy. *Nat Genet* 16: 184-187.

Jay P, Goze C, Marsollier C, Taviaux S, Hardelin J-P, Koopman P, Berta P (1995) The human Sox11 gene: cloning, chromosomal assignment and tissue expression. *Genomics* 29: 541-545.

Jay P, Sahly I, Goze C, Taviaux S, Poulat F, Couly G, Abitbol M, Berta P (1997) Sox22 is a new member of the Sox gene family, mainly expressed in human nervous tissue. *Hum Mol Genet* 6: 1069-1077.

Johnson PF, Sterneck E, Williams SC (1993) Activation domains of transcriptional regulatory proteins. *J Nutr Biochem* 4: 386-398.

Jordan T, Hanson I, Zaletayev D, Hodson S, Prosser J, Seawright A, Hastied N, van Heyningen (1992) The human PAX6 gene is mutated in two patients with aniridia. *Nat Genet* 1: 328-332.

Kainulainen,K, Karttunen L, Puhakka L, Sakai L, Peltonen L (1994) Mutations in the fibrillin gene responsible for dominant ectopia lentis and neonatal Marfan syndrome. *Nat Genet* 6: 64-69.

Kamachi Y, Cheah KSE, Kondoh H (1999). Mechanism of regulatory target selection by the SOX high-mobility-group domain proteins as revealed by comparison of SOX1/2/3 and SOX9. *Mol Cell Biol* 19: 107-120.

Kamachi Y, Kondoh H (1993) Overlapping positive and negative regulatory elements determine lens-specific activity of the d1-crystallin enhancer. *Mol Cell Biol* 13: 5206-5215.

Kamachi Y, Sockanathan S, Liu Q, Breitman M, Lovell-Badge R, Kondoh H (1995) Involvement of SOX proteins in lens-specific activation of crystallin genes. *EMBO Journal* 14: 3510-3519.

Kamachi Y, Uchikawa M, Collignon J, Lovell-Badge R, Kondoh H (1998) Involvement of *Sox1*, *2* and *3* in the early and subsequent molecular events of lens induction. *Development* 125: 2521-2532.

Kamachi Y, Uchikawa M, Kondoh H (2000) Pairing SOX off with partners in the regulation of embryonic development. *Trends Genet* 16: 182-187.

Kanai Y, Kanaiazuma M, Noce T, Saido TC, Shiroishi T, Hayashi Y, Yazaki K (1996) Identification of two Sox17 messenger RNA isoforms, with and without the high mobility group box region and their differential expression in mouse spermatogenesis. *J Cell Biol* 133: 667-681.

Kannabiran C, Wawrousek E, Sergeev Y, Rao GN, Kaiser-Kupfer M, Hejtmancik JF (1999) Mutation of beta-A3/A1-crystallin gene in autosomal dominant zonular cataract with sutural opacities results in a protein with single globular domain. *Invest Ophthalmol Vis Sci* 40: S786.

Kashtan CE (1995) Clinical and molecular diagnosis of Alport syndrome. *Proc Assoc Am Phys* 107: 306-313.

Keller SA, Jones JM, Boyle A, Barrow LL, Killen PD, Green DG, Kapousta NV, Hitchcock PF, Swank RT, Meisler MH (1994) Kidney and retinal defects (Krd), a transgene-induced mutation with a deletion of mouse chromosome 19 that includes the Pax2 locus. *Genomics* 23: 309-320.

Kidson SH, Kume T, Deng K, Winfrey V, Hogan BL (1999) The forkhead/winged-helix gene, Mf1, is necessary for the normal development of the cornea and formation of the anterior chamber in the mouse eye. *Dev Biol (Orlando).211*: 306-22.

Kirby DB (1927) The cultivation of the lens epithelium in vitro. *J Exp Med* 45: 1009-1016.

Kishi M, Mizuseki K, Sasai N, Yamazaki H, Shiota K, Nakanishi S, Sasai Y (2000) Requirement of Sox2-mediated signaling for differentiation of early Xenopus neuroectoderm. *Development* 127: 791-800.

Kjærulff S, Dooijes D, Clevers H, Nielsen O (1997) Cell differentiation by interaction of two HMG-box proteins: Mat1-Mc activates M cell-specific genes in S. pombe by recruiting the ubiquitous transcription factor Ste11 to weak binding sites. *EMBO J* 16: 4021-4033.

Koizumi T, Hendel E, Lalley PA, Tchetgen MB, Nadeau JH (1995) Homologs of genes and anonymous loci on human chromosome 13 map to mouse chromosomes 8 and 14. *Mamm. Genome* 6: 263-268.

Kondoh H (1999) Transcription factors for lens development assessed in vivo. *Curr Opin in Genet Dev* 9: 301–308.

Koopman P (1999) *Sry* and *Sox9*: mammalian testis-determining genes. *Cellular and Molecular Life Sciences* 55: 839-856.

Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R (1991) Male development of chromosomally female mice transgenic for SRY. *Nature* 351: 117-121.

Kozak, M (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nuc Acids Res* 15: 8125-8148.

Krumlauf R (1994) Hox genes in vertebrate development. *Cell* 78: 191-201.

Kuhlbrodt K, Herbarth B, Sock E, Enderich J, Hermans-Borgmeyer I, Wegner M (1998) Cooperative Function of POU Proteins and SOX Proteins in Glial Cells. *J Biol Chem* 273: 16050-16057.

Kuhlbrodt K, Herbarth B, Sock E, Hermans-Borgmeyer I, Wegner M (1998a) *Sox10*, a novel transcriptional modulator in glial cells. *J Neurosci* 18: 237-250.

Kuhlbrodt K, Schmidt C, Sock E, Pingault V, Bondurand N, Goossens M, Wegner M (1998b) Functional analysis of Sox10 mutations found in human Waardenburg-Hirschsprung patients. *J Biol Chem* 273: 23033-23038.

Kulak SC, Kozlowski K, Semina EV, Pearce WG, Walter MA (1998) Mutation in the RIEG1 gene in patients with iridogoniodysgenesis syndrome. *Hum Molec Genet* 7: 1113-1117.

Kume T, Deng K-Y, Winfrey V, Gould, DB, Walter MA, Hogan, BLM (1998) The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic mouse mutation congenital hydrocephalus. *Cell* 93: 985-996.

Kwok C, Weller P, Guioli S, Foster JW, Mansour S, Zuffardi O, Punnett HH, Dominguez-Steglich MA, Brook JD, Young ID, Goodfellow PN, Schafer A (1995) Mutations in SOX 9, the gene responsible for Campomelic dysplasia and autosomal sex reversal. *Am J Hum Genet* 57: 1028-1036.

Laudet V, Stehelin D, Clevers H (1993) Ancestry and diversity of the HMG box superfamily. *Nuc Acids Res* 21: 2493-2501.

Lefebvre V, Huang W, Harley VR, Goodfellow PN, DeCrombrugghe B (1997) Sox9 is a potent activator of the chondrocyte-specific enhancer of the pro[alpha]1(II) collagen gene. *Mol Cell Biol* 17: 2336-2346.

Lefebvre V, Li P, de Crombrugghe B (1998) A new long form of Sox5(L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO J* 17: 5718-5733.

Liegeois NJ, Horner JW, DePinho RA (1996) Lens complementation system for the genetic analysis of growth, differentiation, and apoptosis in vivo. *Proc Natl Acad Sci USA* 93: 1303-1307.

Litt M, Carrero-Valenzuela R, LaMorticella DM, Schultz DW, Mitchell TN, Kramer P, Maumenee IH (1997) Autosomal dominant cerulean cataract is associated with a chain termination mutation in the human beta-crystallin gene CRYBB2. *Hum Molec Genet* 6: 665-668.

Litt M, Kramer P, LaMorticella DM, Murphey W, Lovrien EW, Weleber RG (1998) Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene CRYAA. *Hum Molec Genet* 7: 471-474.

Logan M, Pagan-Westphal SM, Smith DM, Paganessi L, Tabin CJ (1998) The transcription factor Pitx2 mediates situs-specific morphogenesis in response to left-right asymmetric signals. *Cell* 94: 307-317.

Lu MF, Pressman C, Dyer R, Johnson RL, Martin JF (1999) Function of Rieger syndrome gene in left-right asymmetry and craniofacial development. *Nature* 401: 276-278.

MacDonald IM, Haney P, Musarella MA (1998) Summary of ocular genetic disorders and inherited systemic conditions with eye findings. *Ophthal Genet* 19: 1-17.

Mackay D, Ionides A, Kibar Z, Rouleau G, Berry V, Moore A, Shiels A, Bhattacharya S (1999) Connexin46 mutations in autosomal dominant congenital cataract. *Am J Hum Genet* 64: 1357-1364.

Malas S, Duthie S, Deloukas P, Episkopou V (1999) The isolation and high-resolution chromosomal mapping of human SOX14 and SOX21; two members of the SOX gene family related to SOX1, SOX2, and SOX3. *Mamm Genome* 10: 934-937.

Malas S, Duthie S, Mohri F, Lovell-Badge R, Episkopou V (1996) Genetic and physical mapping of the murine *SOX1* gene. *Mamm Genome* 7: 620-621.

Malas S, Sartor M, Duthie S, Hadjantonakis K, Lovell-Badge R, Episkopou V (1997) Cloning and mapping of the human *SOX1*: a highly conserved gene expressed in the developing brain. *Mamm Genome* 8: 866-868.

Mann I (1964) *The development of the human eye.* New York: Grune and Stratton, Inc.

Mansour S, Hall C, Pembrey M, Young I (1995) A clinical and genetic study of campomelic dysplasia. *J Med Genet* 32: 415- 420.

Mansouri A, Goudreau G, Gruss P (1999) Pax genes and their role in organogenesis. *Cancer Res (Suppl.)* 59: 1017s-1710s.

Maroteaux P, Spranger JW, Opitz JM, Kucera J, Lowry RB, Schimke RN, Kagan SM (1971) Le syndrome campomelique. *Presse Med* 22: 1157-1162.

McDowall S, Argentaro A, Ranganathan S, Weller P, Mertin S, Mansour S, Tolmie J, Harley V (1999) Functional and structural studies of wild type SOX9 and mutations causing campomelic dysplasia. *J Biol Chem* 274: 24023-24030.

McElreavey K, Vilain E, Herskowitz I, Fellous M (1993) A regulatory cascade hypothesis for mammalian sex determination: SRY represses a negative regulator of male development. *Proc Natl Acad Sci USA* 90: 3368-3372.

McKusick VA (1992) *Mendelian Inheritance in Man.* Baltimore: The Johns Hopkins Press, 1262-1263.

Mears AJ, Jordan T, Mirzayans F, Dubois S, Kume T, Parlee M, Ritch R, Koop B, Kuo W-L, Collins C, Marshall J, Gould DB, Pearce W, Carlsson P, Enerback S, Morissette J, Bhattacharya S, Hogan B, Raymond V, Walter MA (1998) Mutations of the forkhead/winged-helix gene, FKHL7, in patients with Axenfeld-Rieger anomaly. *Am J Hum Genet* 63: 1316-1328.

Mertin S, McDowall SG, Harley VR (1999) The DNA-binding specificity of SOX9 and other SOX proteins. *Nuc Acids Res* 27: 1359-1364.

Meyer J, Südbeck P, Held M, Wagner T, Schmitz ML, Bricarelli FD, Eggermont E, Friedrich U, Haas OA, Kobelt A, Leroy JG, Van Maldergem L, Michel E, Mitulla B, Pfeiffer RA, Schinzel A, Schmidt H, Scherer G (1997) Mutational analysis of the SOX 9 gene in campomelic dysplasia and autosomal sex-reversal: lack of genotype/phenotype correlations. *Hum Molec Genet* 6: 91-98.

Meyer J, Wirth J, Held M, Schempp W, Scherer G (1996) SOX20, a new member of the SOX gene family, is located on chromosome 17p13. *Cytogenetic Cell Genet* 72: 246-249.

Mirzayans F. Gould DB. Heon E. Billingsley GD. Cheung JC. Mears AJ. Walter MA (2000) Axenfeld-Rieger syndrome resulting from mutation of the FKHL7 gene on chromosome 6p25. *Eur J Hum Genet* 8: 71-4.

Mirzayans F, Pearce WG, MacDonald IM, Walter MA (1995) Mutation of PAX6 gene in patients with autosomal dominant kertitis. *Am J Hum Genet* 57: 539-548.

Mitchell P, Tjian R (1989) Transcription regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 245: 371-378.

MiyataS, MiyashitaK, HosoyamaY (1996) Sry-related genes in Xenopus oocytes. *Biochim Biophys Acta* 1308: 23-27.

Mizuseki K, Kishi M, Matsui M, Nakanishi S, Sasai Y (1998) Xenopus Zic-related-1 and Sox-2, two factors induced by chordin, have distinct activities in the initiation of neural induction. *Development* 125: 579-587.

Moore AT (1992) Cone and cone-rod dystrophies. *J Med Genet* 29: 289-290.

Mumm S, Zucci I, Pilia G (1997) *SOX3* gene maps near DXS984 in Xq27.1, within candidate regions for several X-linked disorders. *Am J Med Gen* 72: 376-378.

Munier FL, Korvatska E, Djemai A, Le Paslier D, Zografos L, Pes[illegible] G, Schorderet DF (1997) Kerato-epithelin mutations in four 5q31-linked corneal dystrophies. *Nat Genet* 15: 247-251.

Muragaki Y, Mundlos S, Upton J, Olsen BR (1996) Altered growth and branching patterns in syndactyly caused by mutations in *HOXD13*. *Science* 272: 548-551.

Nambu PA, Nambu JR (1996) The Drosophila fish-hook gene encodes a HMG domain protein essential for segmentation and CNS development. *Development* 122: 3467-3475.

Ng L-J, Wheatley S, Muscat GEO, Conway-Campbell J, Bowles J, Wright E, Bell DM, Tam PPL, Cheah KSE, Koopman P (1997) SOX9 binds DNA, activates transcription and coexpresses with type II collagen during chondrogenesis in the mouse. *Dev Biol* 183: 108-121.

Nichols BE, Mackey DA, Ritch R, Kalenak JW, Craven ER, Sheffield VC (1997) Identification of a gene that causes primary open angle glaucoma. *Science* 275: 668-670.

Nishiguchi S, Wood H, Kondoh H, Lovell-Badge R, Episkopou,V (1998) Sox1 directly regulates the [gamma]-crystallin genes and is essential for lens development in mice. *Genes Dev* 12: 776-781.

Nishimoto M, Fukushima A, Okuda A, Muramatsu M (1999) The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol Cell Biol* 19: 5453-5465.

Nishimura DY, Swiderski RE, Alward WLM, Searby CC, Patil SR, Bennet SR, Kanis AB, Gastier JM, Stone EM, Sheffield VC (1998) The forkhead transcription factor gene FKHL7 is responsible for glaucoma phenotypes which map to 6p25. *Nat Genet* 19: 140-147.

Nishina S, Kohsaka S, Yamaguchi Y, Handa H, Kawakami A, FujisawaH, Azuma N (1999) PAX6 expression in the developing human eye. *B J Ophthalmol* 83: 723-727.

Ohno S (1967) *Sex Chromosomes and Sex-Linked Genes*. Springer, Berlin.

Osaki E, Nishina Y, Inazawa J, Copeland N, Gilbert D, Jenkins N, Ohsugi M, Tezuka T, Yoshida M, Semba K (1999) Identification of a novel Sry-related gene and its germ cell-specific expression. *Nuc Acids Res* 27: 2503-2510.

Pansky B (1982) *Review of medical embryology*. Toronto: Collier Macmillan Canada, Inc.

Peirano RI, Goerich DE, Riethmacher D, Wegner M (2000) Protein Zero Gene Expression Is Regulated by the Glial Transcription Factor Sox10. *Mol Cell Biol* 20: 3198-3209.

Pellegata NS, Dieguez-Lucena JL, Joensuu T, Lau S, Montgomery KT, Krahe R, Kivela T, Kucherlapati R, Forsius H, de la Chapelle A (2000) Mutations in KERA, encoding keratocan, cause cornea plana. *Nat Genet* 25: 91-95.

Percin EF, Ploder LA, Yu JJ, Arici K, Horsford DJ, Rutherford A, Bapat B, Cox DW, Duncan AMV, Kalnins VI, Kocak-Altintas A, Sowden JC, Traboulsi E, Sarfarazi M, McInnes RR (2000) Human microphthalmia associated with mutations in the retinal homeobox gene CHX10. *Nat Genet* 25: 397-401.

Perrault I, Rozet J-M, Gerber S, Ghazi I, Leowski C, Ducroq D, Souied E, Dufier J-L, Munnich A, Kaplan J (1999) Leber congenital amaurosis. *Molec Genet Metab* 68: 200-208.

Pevny LH, Lovell-Badge R (1997) Sox genes find their feet. *Curr Opin Genet Dev* 7: 338-344.

Pevny LH, Sockanathan S, Placzek M, Lovell-Badge R (1998) A role for SOX1 in neural determination. *Development* 125: 1967-1978.

Pfeifer D, Poulat F, Holinski-Feder E, Kooy F, Scherer G (2000) The SOX8 gene is located within 700 kb of the tip of chromosome 16p and is deleted in a patient with ATR-16 syndrome. *Genomics* 63: 108-116.

Pierrou S, Hellqvist M, Samuelsson L, Enerback S, Carlsson P (1994) Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending. *EMBO J* 13: 5002-5012.

Pingault V, Bondurand N, Kuhlbrodt K, Goerich DE, Prehu M-O, Puliti A, Herbarth B, Hermans-Borgmeyer I, Legius E, Matthijs G, Amiel J, Lyonnet S, Ceccherini I, Romeo G, Smith JC, Read AP, Wegner M, Goossens M (1998) Sox10 mutations in patients with Waardenburg-Hirschsprung disease. *Nat Genet* 18: 171-173.

Plaza S, Dozier C, Saule S (1993) Quail Pax 6 (Pax-QNR) encodes a transcription factor able to bind and trans-activate its own promoter. *Cell Growth Differ* 4: 1041-1050.

Pleasure SJ, Page C, Lee VM-Y (1992) Pure, postmitotic, polarized human neurons derived from NTera 2 cells provide a system for expressing exogenous proteins in terminally differentiated neurons. *J Neurosci* 12: 1802-1815.

Pontiggia A, Rimini R, Harley VR, Goodfellow PN, Lovell-Badge R, Bianchi ME (1994) Sex-reversing mutations affect the architecture of SRY-DNA complexes. *EMBO J* 13: 6115-6124.

Poulat F, de Santa Barbara P, Desclozeaux M, Soullier S, Moniot B, Bonneaud N, Boizet B, Berta P (1997) The human testis determining factor SRY binds a nuclear factor containing PDZ protein interaction domains. *J Biol Chem* 272: 7167-7172.

Poulat,F, Girard F, Chevron M-P, Goze C, Rebillard X, Calas B, Lamb N, Berta P (1995) Nuclear localization of the testis determining gene product SRY. *J Cell Biol* 128: 737-748.

Prior HM, Walter MA (1996) Sox genes: architects of development. *Molecular Medicine* 2: 405-412.

Pusch C, Hustert E, Pfeifer D, Südbeck P, Kist R, Roe B, Wang Z, Balling R, Blin N, Scherer G (1998) The SOX10/Sox10 gene from human and mouse:

sequence, expression and transactivation by the encoded HMG domain transcription factor. *Hum Genet* 103: 115-123.

Quinn JC, West JD, Hill RE (1996) Multiple functions for Pax6 in mouse eye and nasal development. *Genes Dev* 10: 435-46.

Quiring R, Walldorf U, Kloter U, Gehring WJ (1994) Homology of the eyeless gene of Drosophila to the Small eye gene in mice and aniridia in humans. *Science* 265: 785-789.

Read AP, Newton VE (1997) Waardenburg syndrome. *J Med Genet* 34: 656-665.

Ren Z, Li A, Shastry BS, Padma T, Ayyagari R, Scott MH, Parks MM, Kaiser-Kupfer MI, Hejtmancik JF (2000) A 5-base insertion in the gamma-C-crystallin gene is associated with autosomal dominant variable zonular pulverulent cataract. *Hum Genet* 106: 531-537.

Rex M, Orme A, Uwanogho D, Tointon K, Wigmore PM, Sharpe PT, Scotting PJ (1997) Dynamic expression of chicken Sox2 and Sox3 genes in ectoderm induced to form neural tissue. *Dev Dyn* 209: 323-332.

Rex M, Uwanogho DA, Orme A, Scotting PJ, Sharpe PT (1997) cSox21 exhibits a complex and dynamic pattern of transcription during embryonic development of the chick central nervous system. *Mech Dev* 66: 39-53.

Rivera H, Gonzales-Flores SA, Rivas F, Sanchez-Corona J, Moller M, Cantu JM (1984) Monosomy 13q32.3-qter: report of two cases. *Hum Genet* 67: 142-145.

Roebroek AJM, van de Velde HJK, Van Bokhoven A, Broers JLV, Ramaekers FCS, Van de Ven WJM (1993) Cloning and expression of alternative transcripts of a novel neuroendocrine-specific gene and identification of its 135-kDa translational product. *J Biol Chem* 268: 13439-13447.

Roose J, Korver W, de Boer R, Kuipers J, Hurenkamp J, Clevers H (1999) The Sox-13 gene: structure, promoter characterization, and chromosomal localization. *Genomics* 57: 301-305.

Russell SRH, Sanchez Soriano N, Wright CR, Ashburner M (1996) The Dichaete gene of Drosophila melanogaster encodes a SOX-domain protein required for embryonic segmentation. *Development* 122: 3669-3676.

Salvini-Plawen L, Mayr E (1961) in *Evolutionary Biology*, (Vol. 10) (Hecht MK, Steere WC, Wallace B, eds), pp. 207–263, Plenum Press.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: a laboratory manual* (second edition). Cold Spring Harbor Laboratory Press, Plainview.

Sanyanusin P, McNoe LA, Sullivan MJ, Weaver RG, Eccles MR (1995) Mutation of PAX2 in two siblings with renal-coloboma syndrome. *Hum Mol Genet* 4: 2183-2184.

Sanyanusin P, Schimmenti LA, McNoe LA, et al. (1995) Mutation of the PAX2 gene in a family with optic nerve colobomas, renal anomalies and vesicoureteral reflux. *Nat Genet* 9: 358-363.

Schilham MW, Moerer P, Cumano A, Clevers HC (1997) Sox-4 facilitates thymocyte differentiation. *Eur J Immunol* 27: 1292-1295.

Schilham MW, Oosterwegel MA, Moerer P, Ya J, Deboer PAJ, Vandewetering M, Verbeek S, Lamers WH, Kruisbeek AM, Cumano A, Clevers H (1996) Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. *Nature* 380: 711-714.

Schilham MW, Vaneijk M, Vandewetering M, Clevers HC (1993) The murine Sox-4 protein is encoded on a single exon. *Nuc Acids Res* 21: 2009.

Schinzel A (1994) *Human cytogenetics database*. Oxford University Press, Oxford.

Semina EV, Ferrell RE, Mintz-Hittner HA, Bitoun P, Alward WLM, Reiter RS, Funkhauser C, Daack-Hirsch S, Murray,JC (1998) A novel homeobox gene PITX3 is mutated in families with autosomal-dominant cataracts and ASMD. *Nat Genet* 19: 167-170.

Semina EV, Murray JC, Reiter R, Hrstka RF, Graw J (2000) Deletion in the promoter region and altered expression of Pitx3 homeobox gene in aphakia mice *Hum Mol Genet* 9: 1575-1585.

Semina EV, Reiter R, Leysens NJ, Alward WLM, Small KW, Datson NA, Siegel-Bartelt J, Bierke-Nelson D, Bitoun P, Zabel BU, Carey JC, Murray JC (1996) Cloning and characterization of a novel bicoid-related homeobox transcription factor gene, RIEG, involved in Rieger syndrome. *Nat Genet* 14: 392-399.

Shiels A, Mackay D, Ionides A, Berry V, Moore A, Bhattacharya S (1998) A missense mutation in the human connexin50 gene (GJA8) underlies autosomal dominant 'zonular pulverulent' cataract, on chromosome 1q. *Am J Hum Genet* 62: 526-532.

Sinclair A, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346: 240-244.

Smith RS, Zabaleta A, Kume T, Savinova OV, Kidson SH, Martin JE, Nishimura DY, Alward WLM, Hogan BLM, John SWM (2000) Haploinsufficiency of the transcription factors FOXC1 and FOXC2 results in aberrant ocular development. *Hum Molec Genet* 9: 1021-1032.

Southard-Smith EM, Kos L, Pavan WJ (1998) Sox10 mutation disrupts neural crest development in Dom Hirschsprung mouse model. *Nat Genet* 18: 60-64.

Spemann H (1938) *Embryonic development and induction*. New Haven: Yale University Press.

Stephan DA, Gillanders E, Vanderveen D, Freas-Lutz D, Wistow G, Baxevanis AD, Robbins CM, VanAuken A, Quesenberry MI, Bailey-Wilson J, Juo S-HH, Trent J M, Smith L, Brownstein MJ (1999) Progressive juvenile-onset punctate cataracts caused by mutation of the gamma-D-crystallin gene. *Proc Natl Acad Sci USA* 96: 1008-1012.

Stevanovic M, Lovell-Badge R, Collignon J, Goodfellow P (1993) SOX3 is an X-linked gene related to SRY. *Hum Molec Genet* 2: 2013-2018.

Stevanovic M, Zuffardi O, Collignon J, Lovell-Badge R, Goodfellow P (1994) The cDNA sequence and chromosomal location of the human SOX2 gene. *Mamm Genome* 5: 640-642.

Stoilov I, Akarsu AN, Sarfarazi M (1997) Identification of three different truncating mutations in cytochrome P4501B1 (CYP1B1) as the principal cause of primary congenital glaucoma (buphthalmos) in families linked to the GLC3A locus on chromosome 2p21. *Hum Molec Genet* 6: 641-647.

Stone EM, Fingert JH, Alward WLM, Nguyen TD, Polansky JR, Sunden SLF, Nishimura D, Clark AF, Nystuen A, Strachan T, Read AP (1994) PAX genes. *Curr Opin Genet Dev* 4: 427-438.

Strachan T , Read AP (1994) PAX genes. *Curr Opin Genet Dev* 4: 427-438.

Struhl K (1990) Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. *Trends Biochem* 14: 137-140.

Südbeck P, Schmitz M, Baeuerle P, Scherer G (1996) Sex reversal by loss of the C-terminal transactivation domain of human SOX 9. *Nat Genet* 13: 230-232.

Swain A, Narvaez V, Burgoyne P, Camerino G, Lovell-Badge R (1998) Dax1 antagonizes Sry action in mammalian sex determination. *Nature* 391: 761-767.

Tassabehji M, Newton VE, Read AP (1994) Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene. *Nat Genet* 8: 251-255.

Tassabehji M, Read AP, Newton V, Harris R, Balling R, Gruss P, Strachan T (1992) Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. *Nature* 355: 635-636.

Tellier A-L, Amiel J, Salomon R, Jolly D, Delezoide A-L, Auge J, Gubler M-C, Munnich A, Lyonnet S, Antignac C, Vekemans M, Broyer M, Attie-Bitach T (1998) PAX2 expression during early human development and its mutations in renal hypoplasia with or without coloboma. (Abstract) *Am J Hum Genet* 63 (suppl.): A7.

Tjian R, Maniatis T (1994) Transcriptional activation: a complex puzzle with few easy pieces. *Cell* 77: 5-8.

Tommerup N, Schempp W, Meinecke P, Pedersen S, Bolund L, Brandt C, Goodpasture C, Guldberg P, Held KR, Reinwein H, et al. (1993) Assignment of an autosomal sex reversal locus (SRA1) and campomelic dysplasia (CMPD1) to 17q24.3-25.1. *Nat Genet* 4: 170-174.

Ton CCT, Hirvonen H, Miwa H, Weil MM, Monaghan P, Jordan T, van Heyningen V, Hastie ND, Meijers-Heijboer H, Drechsler M, Royer-Pokora B, Collins F, Swaroop A, Strong LC, Saunders GF (1991) Positional cloning and characterization of a paired box- and homeobox-containing gene from the Aniridia region. *Cell* 67: 1059–1074.

Touraine RL, Attie-Bitach T, Manceau E, Korsch E, Sarda P, Pingault V, Encha-Razavi F, Pelet A, Auge J, Nivelon-Chevallier A, Holschneider AM, Munnes M, Doerfler W, Goossens M, Munnich A, Vekemans M, Lyonnet S (2000) Neurological phenotype in Waardenburg syndrome type 4 correlates with novel SOX10 truncating mutations and expression in developing brain. *Am J Hum Genet* 66: 1496-1503.

Travis A, Amsterdam A, Belanger C, Grosschedl R (1991) LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor - enhancer function. *Genes Dev* 5: 880-894.

Triezenberg SJ (1995) Structure and function of transcriptional activation domains. *Curr Opin Genet Dev* 5: 190-196.

Uchikawa M, Kamachi Y, Kondoh H (1999) Two distinct subgroups of Group B *Sox* genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken. *Mech Dev* 84: 103-120.

Uwanogho D, Rex M, Cartwright EJ, Pearl G, Healy C, Scotting PJ, Sharpe PT (1995) Embryonic expression of the chicken Sox2, Sox3, and Sox11 genes suggests an interactive role in neuronal development. *Mech Dev* 49: 23-36.

van de Velde HJK, Roebroek AJM, van Leeuwen FW, Van de Ven WJM (1994) Molecular analysis of expression in rat brain of NSP-A, a novel neuroendocrine-specific protein of the endoplasmic reticulum. *Molec Brain Res* 23: 81-92.

van de Wetering M, Clevers H (1992) Sequence-specific interaction of the HMG box proteins TCF-1 and SRY occurs within the minor groove of a Watson-Crick double helix. *EMBO J* 11: 3039-3044.

van de Wetering M, Oosterwegel M, van Norren K, Clevers H (1993) Sox-4 and Sry-like HMG protein, is a transcriptional activator in lymphocytes. *EMBO J* 12: 3847-3854.

Vandendries ER, Johnson D, Reinke R (1996) *orthodenticle* is required for photoreceptor cell development in the Drosophila eye. *Dev Biol (Orlando)* 173: 243-255.

Vriz S, Lovell-Badge R (1995) The zebrafish Zf-Sox 19 protein: a novel member of the Sox family which reveals highly conserved motifs outside of the DNA-binding domain. *Gene* 153: 275-276.

Wagner T, Wirth J, Meyer J, Zabel B, Held M, Zimmer J, Pasantes J, Bricarelli FD, Keutel J, Hustert E, Wolf U, Tommerup N, Schempp W, Scherer G (1994) Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX 9. *Cell* 79: 1111-1120.

Walther C, Gruss P (1991) Pax-6, a murine paired gene, is expressed in the developing CNS. *Development* 113: 1435-1449.

Watanabe A, Takeda K, Ploplis B, Tachibana, M(1998) Epistatic relationship between Waardenburg syndrome genes MITF and PAX3. *Nat Genet* 18: 283-286.

Waterman M, Jones K (1990) Purification of TCF-1, a T-cell-specific transcription factor that activiates the T-cell receptor c gene ehancer in a context-dependent manner. *New Biol* 2: 621-636.

Wegner M (1999) From head to toes: the multiple facets of Sox proteins. *Nuc Acids Res* 27: 1409-1420.

Werner MH, Huth JR, Gronenborn A, Clore GM (1995) Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell* 81: 705–714.

Wigle JT, Chowdhury K, Gruss P, Oliver G (1999) Prox1 function is crucial for mouse lens-fibre elongation. *Nat Genet* 21: 318-322.

Wilmore HP, Smith MJ, Wilcox SA, Bell KM, Sinclair A (2000) *SOX14* is a candidate gene for limb defects associated with BPES and Möbius syndrome. *Hum Genet* 106: 269-276.

Wirth J, Wagner T, Meyer J, Pfeiffer RA, Tietze H-U, Schempp W, Scherer G (1996) Translocation breakpoints in three patients with campomelic dysplasia and autosomal sex reversal map more than 130 kb from SOX9. *Hum Genet* 97: 186-193.

Wood H, Episkopou V (1999) Comparative expression of the mouse *Sox1*, *Sox2*, and *Sox3* genes from pre-gastrulation to early somite stages. *Mech Dev* 86: 197-201.

Wride M (1996) Cellular and molecular features of lens differentiation: a review of recent advances. *Differentiation* 61: 77-93.

Wright E, Hargrave MR, Christiansen J, Cooper L, Kun J, Evans T, Gangadharan U, Greenfield A, Koopman P (1995) The Sry-related gene Sox9 is expressed during chondrogenesis in mouse embryos. *Nat Genet* 9: 15-20.

Wright EM, Snopek B, Koopman P (1993) Seven new members of the Sox gene family expressed during mouse development. *Nuc Acids Res* 21: 744.

Wunderle VM, Critcher R, Ashworth A, Goodfellow PN (1996) Cloning and characterization of SOX5, a new member of the human SOX gene family. *Genomics* 36: 354-358.

Wunderle VM, Critcher R, Hastie N, Goodfellow PN, Schedl A (1998) Deletion of longe-range regulatory elements upstream of SOX9 causes campomelic dysplasia. *Proc Natl Acad Sci USA* 95: 10646-10654.

Xu P-X, Zhang X, Heaney S, Yoon A, Michelson AM, Maas RL (1999) Regulation of Pax6 expression is conserved between mice and flies. *Development* 126: 383-395.

Xu P-X, Woo I, Her H, Beier DR, Maas RL (1997) Mouse Eya homologues of the Drosophila eyes absent gene require Pax6 for expression in lens and nasal placode. *Development* 124: 219-231.

Yoshioka H, Meno C, Koshiba K, Sugihara,M, Itoh H, Ishimura Y, Inoue T, Ohuchi H, Semina E, Murray JC, Hamada H, Noji, S (1998) Pitx2, a bicoid-type homeobox gene, is involved in a lefty-signaling pathway in determination of left-right asymmetry. *Cell* 94: 299-305.

Yuan H, Corbi N, Basilico C, Dailey L (1995) Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* 9: 2635-2645.

Zhang K, Bither PP, Park R, Donoso LA, Seidman JG, Seidman CE (1994) A dominant Stargardt's macular dystrophy locus maps to chromosome 13q34. *Arch Ophthalmol* 112: 759-764.

Zhou G, Lefebvre V, Zhang Z, Eberspaecher H, de Crombrugghe B (1998) Three high-mobility group-like sequences within a 48-base pair enhancer of the Col2a1 gene are required for cartilage-specific expression in vivo. *J Biol Chem* 273: 14989-14997.

Zinovieva RD, Duncan MK, Johnson TR, Torres R, Polymeropoulos MH, Tomarev SI (1996) Structure and chromosomal localization of the human homeobox gene Prox 1. *Genomics* 35: 517-522.

APPENDIX A: Human *SOX* Gene Accession Numbers

Gene	Accession Number	Comment
SRY	L08063	
	L10102	
	L10101	
	X53772	
SOX1	Y13436	
SOX2	Z31560	Partial
SOX3	X71135	
SOX4	X70683	
	AF124147	5' UTR
SOX5	NM_006940	
	S83308	Alt. spliced
SOX6	X65663	HMG box
SOX8	NM_014587	
SOX9	Z46629	
	S74504	
	S74505	
	S74506	
	AB022194	Exon 1, promoter
SOX10	NM_006941	
	AJ001183	
SOX11	U23752	
SOX12	X73039	HMG box
SOX13	NM_005686	
	AF083105	
SOX14	AJ006230	
	NM_004189	
SOX18	NM_018419	
SOX20	NM_006942	
SOX21	NM_007084	
SOX22	NM_006943	
SOX25	AF032449	HMG box
SOX26	AF032450	HMG box
SOX27	AF032452	HMG box
SOX28	AF032453	HMG box
SOX29	AF032454	HMG box
SOX30	NM_7017	
SOXL	NM_006339	

APPENDIX B: Solutions

BAAD solution
150 μL BNN buffer
30 μL BAAD (0.2 μg/μL)
25 μL goat serum
295 μL H_2O
500 μL total
*Pre-BAAD is BAAD solution without BAAD added.

BNN buffer
10 mL 1 M $NaHCO_3$
10 mL 0.2 % NaN_3
1 mL 5 % NP-40
9 mL H_2O
30 mL total

Church & Gilbert hybridization solution
24.3 g Na_2HPO_4
10.9 g $NaH_2PO_4 \cdot H_2O$
35.0 g SDS
5.0 g BSA
ddH_2O to 500 mL total volume

DAPI/PI solution
5 mL P-phenylenediamine(10 mg/mL in 0.01 M PBS, pH 8.0)
40 μL DAPI (1 mg/mL)
20 μL PI (1 mg/mL)
45 mL glycerol
Adjust to pH 8.0 with 0.5 M sodium carbonate-bicarbonate buffer.

Denaturing solution
0.5 M NaOH
1.5 M NaCl

DEPC H_2O
0.1% DEPC (1 mL/L)
Stir O/N, then autoclave.

FAD solution
150 μL BNN buffer
6 μL FAD (0.4 μg/μL)
250 μL 10% powdered milk
94 μL H_2O
500 μL total
*Pre-FAD is FAD solution without FAD added.

20X GTB
54 g Tris base
18 g taurine
1 g $Na_2EDTA\ 2H_2O$
ddH_2O to 250 mL total volume

High TE buffer
50 mL 1 M Tris-Cl
(C_f = 100 mM)
40 mL 0.5 M EDTA pH 8.0
(C_f = 40 mM)
ddH2O to 500 mL

GTE
1.25 mL 2 M D-glucose
(C_f = 50 mM)
1 mL 0.5 M EDTA (C_f = 10 mM)
1.25 mL 1 M Tris-HCl pH 8.0
(C_f = 25 mM)
46.5 mL ddH_2O

LB*
6 g bacto-tryptone
3 g yeast extract
6 g NaCl
600 mL ddH_2O
Stir to partially dissolve solutes.
Autoclave.
*For LB agar, add 9 g bactoagar

10X MOPS
46.3 g
3-morpholinopropanesulphonic acid
(C_f = 0.2 M)
16 mL 3 M NaOAc (C_f = 50 mM)
20 mL 0.5 M EDTA (C_f = 10 mM)
ddH_2O to 1 L, pH to 7.0 with NaOH
Autoclave.

Neutralizing solution
0.5 M Tris-Cl pH 7.5
1.5 M NaCl

PE hybridization buffer (10X)
300 μL 5M KCl (C_f = 1.5 M)
100μL 1M Tris-Cl pH 8.4
(C_f = 0.1M)
20 μL 0.5 M EDTA(C_f = 10 mM)
580 μL DEPC H_2O
1000 μL total

10X TBE

3M Potassium acetate (KAc) solution
60 mL 5M KAc (pH 5)
11.5 mL glacial acetic acid
28.5 mL ddH_2O

RBC lysis buffer
4.12 g Tris-base
14.9 g ammonium chloride
dd H_2O to 2000 mL, pH to 7.65

RNase reaction mix
50 μL 10 mg/mL salmon sperm
DNA (C_f = 100 μg/mL)
2 μL 10 mg/mL RNase A
(C_f = 20 μg/mL)
4.95 mL TEN 100 buffer (see below)
5 mL total

20X SSC
175.3 g NaCl
88.2 g trisodium citrate
ddH_2O to 1 L, pH to 7.0 with NaOH

0.5 M Sodium carbonate-bicarbonate buffer
19.2 mL $NaHCO_3$ (8.4 g/100 mL)
1.94 mL Na_2CH_3 (1.06 g/10mL)
28.86 mL H_2O
50 mL total, pH 9.0

SM buffer
2.9 g NaCl
1 g $MgSO_4$
25 mL 1 M Tris-Cl pH 9.0
2.5 mL gelatin (C_f = 2%)
472 mL ddH_2O

108 g Tris base
55 g boric acid
40 mL 0.5 M EDTA
ddH_2O to 1 L total volume

TE buffer
1 mL 1 M Tris-Cl pH 8.0 (C_f = 10 mM)
0.2 mL 0.5 M EDTA (C_f = 1 mM)
98.8 mL ddH_2O

TEN 100 buffer
2 mL 5 M NaCl (C_f = 100 mM)
1 mL 1 M Tris-Cl pH 8.0 (C_f = 10 mM)
0.2 mL 0.5 M EDTA (C_f = 1 mM)
96.8 mL ddH_2O

TENS buffer
250 μL 10% SDS (C_f = 0.5%)
50 μL 10 M NaOH (C_f = 0.1 M)
4.7 mL TE

WBC lysis buffer
40 mL 0.5 M EDTA pH 8.0
50 mL 1 M Tris-Cl pH 8.0
100 mL 5 M NaCl
ddH_2O to 500 mL, pH to 8.0
Autoclave, then add 5 mL 20% SDS.

APPENDIX C: *SOX1* Primers

F/R	Name (Position)	Sequence	Tm	PCR pair (An. Temp.)
F	900-A (-973)	CAG GAA GCG GTG TGG CTG TC	-	-
R	900-D (-954)	GAC AGC CAC ACC GCT TCC	62.9	-
F	5-68 (-931)	CCC TCC TTC CTG CTT TGC	60.9	5-1035 (56*)
R	900-C (-910)	GGG AGC AAA GCA GGA AGG	60.9	-
R	64 (-883)	GGA CCG GGG AGG AGG GGA GGA AGG	75	-
F	5-124 (-875)	GCG CCC AGT GTA TCT ACT CC	59.7	5-1035 (56*)
R	100 (-847)	CGT GGG GAG GGA GTA AAT ACA CTG G	67	-
F	5-345 (-654)	GGC TCT CCG GCC CTC TCC	66.9	-
F	5-345G (-654)	GGC GCG CCG GCC CTC TCC	-	-
R	R-179 (-623)	GAG CGC GGG CTC GCG GAG AGG	78.1	-
R	900-B (-577)	GGG GCG ATT ACG GTG CAG	64.8	-
R	R-364 (-436)	GGT CCT CAG CGG CCT CCA ACT CG	74.1	-
F	12 (-431)	GCA GGA GGA AAG GAG ACA GC	61.5	473 (59) P-221 (69) P-279 (65)
R	R-377 (-423)	TCC TCC TGC GCT CGG TCC TCA GC	75.0	-
R	P-221 (-223)	CAC AAA CCA CTT GCC AAA GAG	60.7	12 (69)
R	P-279 (-165)	GCA GAA ACA CAC GCA CTC GG	61.7	12 (65)
F	311 (-133)	CCA TTC TTC TCT CCG CTA GG	59.0	-
F	351 (-93)	CTC CGT CTG AAT TCC TCT CC	58.8	-
F	355[S] (-87)	CTG AAT TCC TCT CCG TCT CC	58.8	631 (55, 5%)

F/R	Name (Position)	Sequence	Tm	PCR pair (An. Temp.)
R	946 (-30)	GAG GAG AGG GCC TGG AGC ATA GAC G	71.1	-
R	473[S] (18)	CAT CAT CAT GCT GTA CAT CGG	60.0	12 (59)
F	mSOX1-D (19)	GAG ACC GAC CTG CAC TCG	61	-
R	5-1035 (35)	GAG TGC AGG TCG GCT TCC	59.3	5-68 (56*) 5-124 (56*)
R	R-1012 (36)	CGA GTG CAG GTC GGT CTC CAT CAT C	72.1	-
F	mSOX1-F[S] (130)	ACC AAG GCC AAC CAG GAT C	62	mSOX1-G (58)
F	582[S] (138)	CAA CCA GGA CCG GGT CAA	63.0	857 (60, 5%)
R	5-1148 (148)	GGT CCT GGT TGG CCT TGG	63.8	-
F	DEG-1(F) (154)	AAG CGA CCC ATG AAY GCN TT	-	-
R	5-1168 (168)	GTT CAT GGG CCG TTT GAC C	63.6	-
R	631[S] (187)	GGG ACC ACA CCA TGA AGG	59.7	355 (55, 5%)
R	5-1219 (219)	GTT CTC CTG GGC CAT CTT GC	64.3	-
R	661 (246)	TCT CCT GGG CCA TCT TGC	62.3	-
R	1259 (283)	CCT CGG ACA TGA CCT TCC ACT CG	69.3	-
F	792 (343)	CAC C_G GAT TAC AAG TAC GGC	60	-
F	795[S] (343)	CAC CCG GAT TAC AAG TAC CG	60.2	-
R	DEG-1(R) (360)	GTA CTT RTA RTY NGG RTA	-	-
F	SRY-5' (369)	CCG CAA GAC CAA GAC GCT G	65	-
F	820[S] (376)	ACC AAG ACG CTG CTC AAG	57.6	1181 (62, 5%)
R	*mSOX1*-G[S] (404)	TAC TTG TCC TTC TTG AGC AGC G	62	MSOX1-F (58)
R	857[S] (413)	GCC AGC GAG TAC TTG TCC TTC	59.5	582 (60, 5%)

F/R	Name (Position)	Sequence	Tm	PCR pair (An. Temp.)
F	1043 ^{S} (609)	GGC CAT GAT GCA GGA GGC	65.2	1241 (60, 5%) 1473 (66)
R	mSOX1-H (726)	GTT GTG AGG ATG CGC G	58	-
F	1180 ^{S} (736)	ATG CAC CGC TAC GAC ATG G	63.1	1473 (56, 5%*)
R	1181 ^{S} (737)	ATG GGC TGC GGG TTG TGC	67.9	820 (62, 5%)
R	PA-3' (737)	ATG GGC TGC GGG TTG TGA G	67	-
R	1241 ^{S} (797)	ATG TAG CCC TGC GAG TTG G	61.2	1043 (60, 5%)
F	mSOX1-A (817)	TAC GGC GGC ATC CCT TAC G	-	-
F	1384 ^{S} (940)	CTG GGC TCT CTG GTG AAG TC	60.0	MoMo2 (58, 5%)
R	1398 (958)	ACT TCA CCA GAG AGC CCA GC	62	-
R	1398d (958)	ACT T_C CAG AGA GCC CAG C	-	-
R	mSOX1-C (967)	TCG GCT CCG ACT TCA CCA G	-	-
R	1473 ^{S} (1039)	ACA TGC TGA TCA TCT CGC G	61.0	1043 (66) 1180(56, 5%*)
R	mSOX1-E (1042)	GGT ACA TGC TGA TCA TCT CGC	61	-
F	R-2138 (1162)	CTG ACG CAC ATC TAG CGC CTT CG	70.3	-
F	359F (1165)	ACG CAC ATC TAG CGC CTT CG	65.9	359R (66, 5%)
F	1611 (1168)	CAC ATC TAG CGC CTT CGG	60.5	2077 (60, 7.5%*)
F	MoMo1 (1170)	CAT CTA GCG CCT TCG CC	60	-
R	1632 ^{S} (1184)	CGA AGG CGC TAG ATG TGC	60.7	-
R	MoMo2 ^{S} (1294)	TGA CAA GAA TGT GGG AAC G	58.0	1384 (58, 5%)
F	2233 (1358)	GGC GTG GCT TTT GTA CAG ACG TTC C	70.4	-

F/R	Name (Position)	Sequence	Tm	PCR pair (An. Temp.)
F	1809 (1375)	GCT GAC ACC AGA CTT GGG	58	1993 (58*)
R	1836 (1394)	AAC CCA AGT CTG GTG TCA GC	60.2	-
F	1889[S] (1447)	TAC GGA GGA AGC GGA AAG CG	67	2077 (63/61 TD)
R	359R (1523)	AAG TGG GCT TCG CCT CTC G	65.1	359F (66, 5%)
R	1993 (1524)	AAA GTC GTG TCG CCT CTC G	62	1809 (58*)
R	mSOX1-B (1542)	AGA GCC GGC AGT CAT ACA A	-	-
F	1972 (1546)	ACT TTC CTC CGC GTT GC	60	-
R	2046 (1620)	ATA ACA AAT GCG ACG ACG TCC G	66	-
R	2077 (1651)	TGT AAG TGC TTG GAC CTG CC	61.2	1611 (60, 7.5%*)
R	2149 (1723)	GGG TGG TGG TGG TAA TGT G	60	-
F	506F (1731)	CAG ACG GCG GAG TTA TAT TCT GG	64.1	506R (66)
F	367F (1945)	CGA GTC AAC AAA TTT AAG AGA CAG AGC	62.9	367R (66)
F	336F (1976)	TTT CTC CAT AAA TTT GTA ACA TGC	57.4	-
R	506R (2236)	CCG TCC GCA GTG AAA CAG C	65.3	506F (66)
R	367R (2311)	ACC AAA CTG TGG CGC AGA CC	65.8	367F (66)
R	336R (2311)	ACC AAA CTG TGG CGC AGA CC	65.8	-
F	FB8-1086 (2461)	AAC ACA _TC GCT GAA _C_ AAA CG	59.8	-

***SOX1* Primers Legend:**

F/R forward or reverse direction

Name primer name
a superscript "S" follows the name of primers used for sequencing
names preceeded by 5- indicate primers designed to detect a 5' intron
names preceeded by R- indicate primers designed for RACE
names preceeded by P- indicate primers designed for primer extension

Position location of first base of primer (position 1 is A of the start ATG)

Sequence primer sequence listed 5' to 3'
underlined bases or spaces do not match *SOX1* sequence

Tm estimated melting temperature based on Primer3 or Geneworks™
a dash (-) indicates no information available

PCR pair primer pair for PCR amplification
a dash (-) indicates no information available

An. Temp. PCR annealing temperature(°C) and amount DMSO (%)
TD indicates touchdown PCR

* PCR cycling times are 30 sec. denaturing, 30 sec. annealing, and 30 sec. extension with the following exceptions:
1384 + MoMo2: denature 45 sec., anneal 30 sec., extend 60 sec.
1611 + 2077: denature 30 sec, anneal 30 sec., extend 60 sec.
1809 + 1993: denature 30 sec, anneal 30 sec., extend 60 sec.
5-68 + 5-1035: denature 30 sec, anneal 30 sec., extend 90 sec., with G/C rich PCR additive (Clontech)
5-124 + 5-1035: denature 30 sec, anneal 30 sec., extend 90 sec., with G/C rich PCR additive (Clontech)

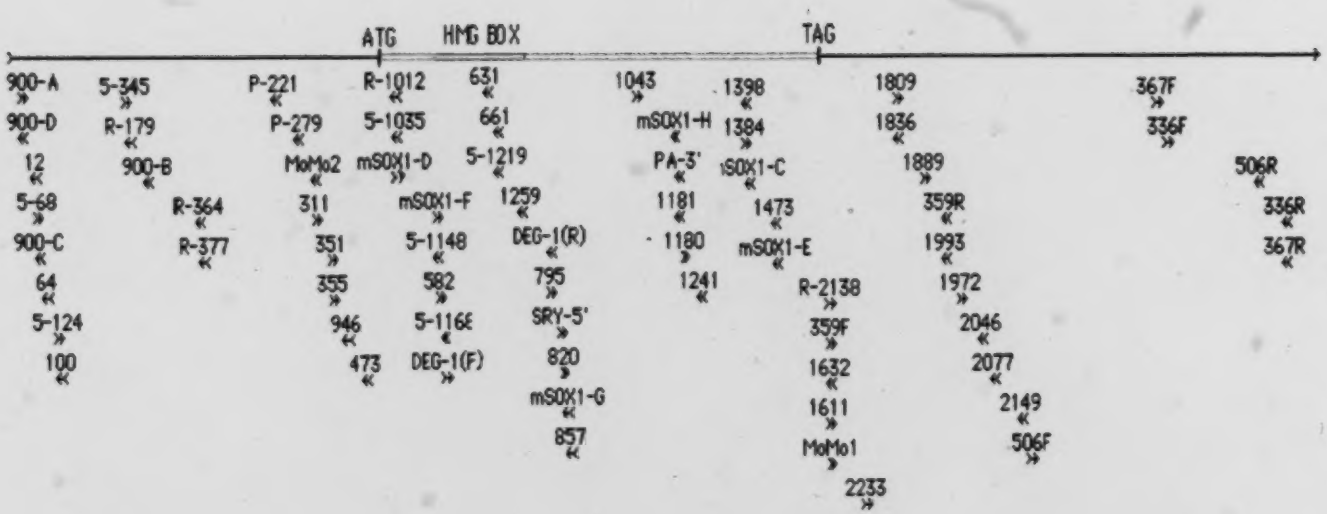



Figure 40. SOX1 PRIMER MAP